SYSTEM ARCHITECTURES AND DIGITAL SIGNAL PROCESSING ALGORITHMS FOR ENHANCING THE OUTPUT AUDIO QUALITY OF STEREO FM BROADCAST RECEIVERS

Juin-Hwey Chen, Thomas Baker, Evan McCarthy, and Jes Thyssen

Broadcom Corporation
Irvine, California, USA

ABSTRACT

This paper presents two FM receiver architectures and three digital signal processing algorithms for enhancing the output audio quality of FM broadcast receivers. The two receiver architectures differ only in the front-end processing for estimating the carrier-to-noise ratio and noise floors of the stereo audio signals. The shared back-end processing consists of three algorithms to suppress the noise in the audio signal, to detect and cancel noise pulses (static), and to conceal the degrading effects of fast fading, respectively. Together these FM enhancement techniques achieve about 20 to 35 dB improvements in SNR and stereo separation over a wide range of RF signal strength spanning nearly 30 dB. Perceptually, the audio quality improvement is large and obvious when the received FM signal is weak.

Index Terms— FM enhancement, FM noise reduction.

1. INTRODUCTION

The current stereo frequency modulation (FM) radio broadcasting method has been in use since the early 1960s and is still being used widely throughout the world. When the received FM radio signal is weak, there are at least four possible quality-degrading effects:

(1) increased level of hissing noise,
(2) reduced stereo separation to keep the noise low,
(3) appearance of noise pulses ("static") when the FM radio signal strength is below a threshold, and
(4) short but loud bursts of noise if the FM receiver is traveling through deep fades at highway speed.

Researchers have previously proposed techniques to address some of these issues. Al-Nuaimi used a threshold amplifier to blank out noisy intermediate frequency (IF) signals below a threshold and reported improved signal-to-noise ratio (SNR) [1]. Pettigrew and Moir reported a 10 dB reduction of impulse noise and a 6 dB reduction of white noise by combining the outputs of an amplitude-locked loop and a phase-locked loop to get an estimate of the FM noise, and then subtracting it from the noisy FM signal [2]. Lee proposed comparing the output signals from three parallel FM tuners to assemble the best available audio signal and reported that over 95% of noise was eliminated without degrading audio bandwidth or stereo separation [3]. Taura et al. proposed a digital architecture to detect and remove impulse noise, and they also used a stereo reception control process to reduce the multipath noise [4]. Although not addressing stereo FM broadcast specifically, Lukin et al. proposed an adaptive censored estimator for use with a robust de-noising scheme based on Discrete Fourier Transform (DFT) for noisy FM signals, and they reported better results than using other estimators [5].

This paper presents two alternative FM receiver architectures and three digital signal processing algorithms [6] designed to address all four quality-degrading effects listed earlier. DFT-based single-channel noise suppression [7] is used to address the effects (1) and (2). A linear-prediction-based noise pulse detector and a canceller are used to address the effect (3). A waveform extrapolation scheme is used to address the effect (4). These algorithms differ from the previous methods above, and together they achieve large improvements in SNR and stereo separation.

This paper is organized as follows: Section 2 gives an overview of the FM broadcasting system. Section 3 introduces two alternative FM receiver architectures Sections 4 through 6 describe the three back-end signal processing algorithms. Section 7 summarizes the performance and complexity of the proposed FM receivers. Section 8 provides concluding remarks.

2. OVERVIEW OF FM BROADCASTING

In FM broadcasting, the Left (L) and Right (R) channels of a stereo audio signal are transmitted as L+R and L-R signals carrying mono and stereo information, respectively. The L+R signal is transmitted as baseband audio between 30 Hz and 15 kHz while the L-R signal is modulated onto a 38 kHz double-sideband suppressed carrier signal and occupies 23 to 53 kHz. A 19 kHz pilot tone is transmitted at 8–10% of the overall modulation level and used by the FM receiver to regenerate the 38 kHz subcarrier with the correct phase.

A stereo FM receiver demodulates the L+R and L-R signals, and then adds L+R to L-R to recover the L signal and subtracts L-R from L+R to recover the R signal. The SNR of L-R is about 20 dB lower than that of L+R. As the carrier-to-noise ratio (CNR) decreases, the L-R noise dominates the perceived noise in the final L and R signals. Conventional FM receivers use "stereo blending" to reduce the L-R signal progressively with decreasing CNR to reduce the noise in the output signals at the price of reduced stereo separation. As the CNR continues to decrease, eventually the FM receiver output collapses to only the L+R, or mono, signal. When the CNR crosses below 12 dB, an impulsive type of noise appears and is perceived as "static."

When the FM radio signal is subject to multipath distortion and blocking, there will be local areas of low FM signal strengths. If an FM receiver is moving through such areas at highway speed, there will be frequent and brief drops in the CNR to very low levels. This is referred to as "fast fading" in this paper, and it can cause frequent, short, but loud noise bursts in the FM receiver output signal.

3. TWO FM RECEIVER ARCHITECTURES

Two FM receiver architectures have been studied in the context of our proposed FM enhancement (FME) system. Fig. 1 shows the first such architecture. Besides producing the L+R and L-R signals sampled at 32 kHz, the FM demodulator also produces a 32 kHz quadrature-demodulated L-R (q-d L-R) signal by demodulating the L-R signal with a special carrier that is 90 degrees out of phase with the carrier used to demodulate the L-R signal. This special carrier will not pick up the L-R signal due to the orthogonality and will instead produce a quadrature-demodulated version of the L-R noise. Such q-d L-R noise will have roughly the same magnitude spectrum as the true noise in the demodulated L-R signal. Thus, the L-R noise floor, which is a function of frequency, can be estimated by computing the magnitude spectrum of this q-d L-R noise.

The L-R and L+R noise floors move up or down as the CNR changes. Thus, the power of the q-d L-R noise can be used to estimate the CNR through an empirically-derived mapping. Also, at any given CNR, there is a fixed relationship between L+R and L-R noise floors. Hence, after the L-R noise floor is determined, the L+R noise floor is estimated by table lookup and interpolation along the frequency and CNR axes using empirically-derived tables.

The second architecture uses the same back-end processing and differs only in its front-end processing, which is shown in Fig. 2. It is useful for avoiding the hardware needed for producing the q-d L-R noise, and it can be used to retrofit older FM receiver chips with FME.

In Fig. 2, the FM demodulator produces the L+R and L-R signals sampled at 48 kHz rather than 32 kHz. The stop-band noise extractor extracts the 19.4 to 21.4 kHz stop band of the 48 kHz L+R signal by bandpass filtering. This frequency range is chosen to avoid the 19 kHz pilot tone and other spurious spectral peaks. The power of the extracted stop-band noise is calculated and mapped to an estimated CNR through an empirically-derived mapping.

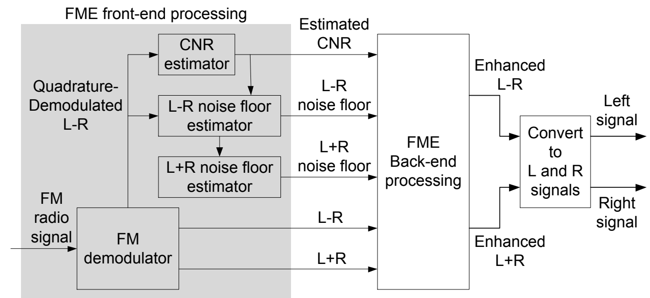

Figure 1 First FM enhancement system architecture

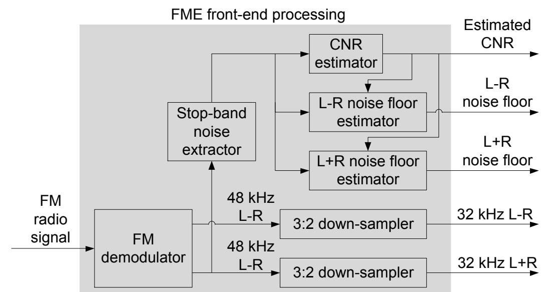

Figure 2 Front-end processing of the second FM enhancement system architecture

The L-R noise floor has an observable relationship with the stop-band noise power and, thus, can be estimated from the latter through an empirically-derived table and interpolation along the frequency and CNR axes. The L+R noise floor can be estimated in a similar fashion but with a different table. The 48 kHz sampled L+R and L-R signals are up-sampled to 96 kHz, then down-sampled to 32 kHz.

Fig. 3 shows the back-end processing. The 32 kHz L+R signal is first processed by L+R pulse suppression (PS) to suppress the impulse noise when CNR is below about 12 dB. The L-R signal does not need PS because at such a low CNR the L-R noise suppression (NS) is so aggressive that the L-R noise pulses are reduced to a negligible magnitude.

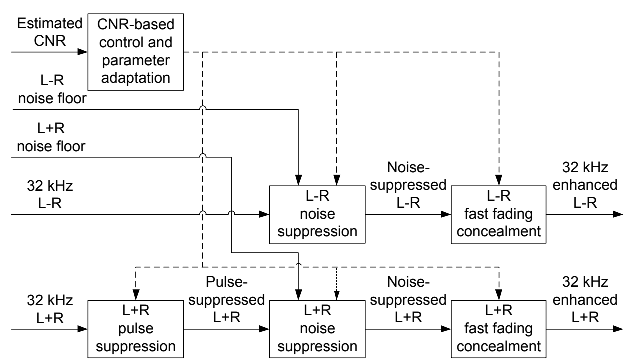

Figure 3 Back-end processing of FM enhancement systems

The L-R signal and pulse-suppressed L+R signal are each processed by an NS stage. The noise-suppressed L-R and L+R signals are processed by fast fading concealment

(FFC) to produce the enhanced L-R and L+R signals, which are then converted to L and R signals as shown in Fig. 1.

At a high CNR when the FM radio signal is strong, PS and FFC are bypassed and NS is turned off or provides very little noise suppression to avoid degrading audio quality. As the CNR decreases, NS increases the degree of noise suppression to keep the output audio signal noise-free and, at some point, the PS and FFC are also turned on. Thus, over a wide range of FM radio signal levels (60 dB or more), the operations of PS, NS, and FFC are controlled by the CNR-based control and parameter adaptation block in Fig. 3.

4. PULSE SUPPRESSION

Noise pulses that are perceived as "pops" or "static" start to appear in the demodulated L+R and L-R signals when the CNR is below 12 dB during silent pauses when no audio is being broadcast. During louder audio broadcasts, this CNR threshold can occasionally go up to about 18 dB.

It is observed that, although multiple noise pulses can sometimes overlap in time, all isolated noise pulses have very similar waveform shapes and durations, and the peak magnitudes of the pulses are within a limited range. Thus, it is possible to build a pulse shape "codebook" to detect such noise pulses and then subtract them from the audio signal to eliminate the pulses. This is called pulse suppression (PS).

To build the "templates" in the pulse shape codebook, 32 different noise pulses with different peak magnitudes and signs were extracted from demodulated L+R signals. They were then up-sampled to 192 kHz, time aligned, normalized, and averaged. The result is the averaged prototype noise pulse shown in Fig. 4. This 192 kHz prototype pulse is then sub-sampled in six different phases to get six pulse templates at 32 kHz. Each template has seven samples.

These pulse templates are used to do template matching in the observed peak magnitude range to detect and remove noise pulses. This is easy for silent regions of the audio signal but more difficult when the pulses are superimposed on an active audio signal such as music or speech.

The PS algorithm uses linear prediction to remove the audio signal as much as possible before attempting to detect noise pulses. It first performs adaptive short-term linear prediction [8] to obtain the prediction residual signal that has the short-term predictable portion of the audio signal removed. It then performs long-term linear prediction [8] on the short-term prediction residual signal to remove the long-term predictable portion of the audio signal.

The final prediction residual signal is then used to do template matching using the "convolved" pulse templates, which are obtained by convolving the pulse templates with the impulse response of the short-term prediction error filter of the current frame. This is because when examined locally, a noise pulse embedded in the audio signal would have been filtered by the short-term prediction error filter when the short-term prediction residual is calculated.

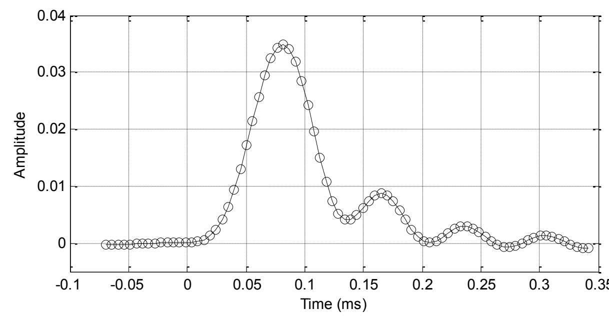

Figure 4 Prototype noise pulse at 192 kHz

The PS algorithm picks seven consecutive samples at a time through the current frame of the final prediction residual signal and compares each extracted 7-dimensional residual vector with the six convolved noise templates. The best-matching template that minimizes the matching error is identified, and if the minimum matching error is smaller than a threshold and the peak magnitude of the residual vector is within the observed pulse magnitude range, then a noise pulse is considered detected, and the best-matching "non-convolved" version of the noise pulse template is subtracted from the L+R signal.

The PS algorithm then subtracts the best-matching convolved pulse template from the short-term prediction residual to prevent the long-term prediction error filter from repeating this detected pulse later. Furthermore, the PS algorithm goes back 3 to 4 samples to look for another possible noise pulse, because if two noise pulses overlap but are separated by at least 3 samples, such sequential detection/removal can eliminate both pulses.

5. DFT-BASED NOISE SUPPRESSION

The problem of reducing hissing noise and enhancing stereo separation can be treated as a single-channel noise suppression (NS) problem. By selectively applying more attenuation at frequencies where the L-R signal is noisier, NS does not apply a small weight across all frequencies as in conventional stereo blending. Thus, not only can NS reduce the noise in the L-R signal (and, eventually, in the L and R signals) more effectively than stereo blending, but also the stereo separation can be better enhanced at the same time as a by-product. Less aggressive NS is also applied to the L+R signal since it also has noise, albeit at a lower level.

Reliable estimation of the noise floor is a prerequisite for successful single-channel NS. In FM applications, the audio signal is often a music signal without any gap that contains only background noise. This makes it difficult to estimate the noise floor reliably by analyzing the audio signal directly. Thus, the front-end processing of the two FME system architectures rely on other means to estimate the L-R and L+R noise floors more reliably.

The NS algorithm in the FME system converts the 32 kHz L-R and L+R signals to the frequency domain by taking 512-point Fast Fourier Transform (FFT) after applying a sine window. Adjacent sine windows are overlapped by 50%, giving a frame size of 256 samples, or 8 ms. With the L-R and L+R noise floors provided by the front-end processing, the instantaneous signal-to-noise ratio (SNR) at each frequency bin is obtained by dividing the signal power by the power spectrum of the noise floor. This SNR is smoothed along the time axis using a first-order all-pole low-pass filter, except that it is not smoothed where the signal is deemed to be active, defined as frequency bins with the instantaneous SNR greater than a threshold.

The smoothed SNR is converted to the "a priori SNR" by subtracting 1 from it and lower-bounding it with a small positive number. The a priori SNR, denoted as $SNR(k)$ for the k-th frequency bin, is mapped to a noise suppression gain $H(k)$ according to the following equation:

$$H(k) = \frac{\alpha * SNR(k) + (1-\alpha) * Hs(k)}{\alpha * SNR(k) + (1-\alpha)}$$

where $Hs(k)$ is the desired noise attenuation for the k-th frequency bin, and α is a parameter between 0 and 1 that controls the trade-off between signal distortion and unnaturalness of the residual noise after NS. The derivation of this equation can be found in [9].

After multiplying the DFT coefficients by the noise suppression gain $H(k)$, an inverse FFT (IFFT) is applied. The result is multiplied by a sine window again, and an overlap-add operation between two adjacent sine windows is performed to reconstruct the L-R and L+R signals.

6. FAST FADING CONCEALMENT

The fast fading concealment (FFC) algorithm monitors instantaneous CNR to detect short, deep fades where the CNR dips to very low levels. When detected, FFC extrapolates previous L+R and L-R waveforms to the current frame using periodic waveform extrapolation, similar to the packet loss concealment techniques described in [10].

If the instantaneous CNR is below the detection threshold but above a mixing threshold, FFC generates a weighted sum of the extrapolated waveform and the normally demodulated and FME-enhanced waveform, with the weights controlled by the instantaneous CNR. If the instantaneous CNR is below the mixing threshold, the normally demodulated signal is too noisy, so only the extrapolated waveform is used. If the number of consecutive frames in a deep fade exceeds a threshold, FFC gradually ramps down the output audio signal level.

7. PERFORMANCE AND COMPLEXITY

At CNR above 9 dB, PS removes more than 99.5% of noise pulses, and the few escaped pulses are often masked by the audio signal and are not audible. Perceptually, PS cleans up static-corrupted audio signals dramatically and makes them sound much cleaner; it extends the listenable range of the RF signal level down by 3 to 3.5 dB. Similarly, NS eliminates the hissing noise and enhances stereo separation very significantly, and FFC greatly improves the perceptual audio quality by concealing short noise bursts.

Figs. 5 and 6 show the SNR and stereo separation, respectively, as functions of the FM signal level. They were objectively measured using real-time recorded tone bursts separated by silence, stepping from an RF signal level of –60 dBm to –115 dBm. Large improvements of 20 to 35 dB are achieved over a nearly 30 dB range of RF signal level.

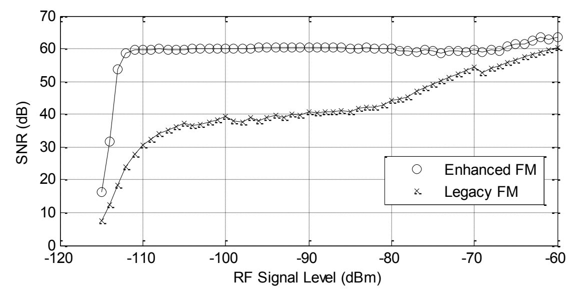

Figure 5 SNR versus RF signal level

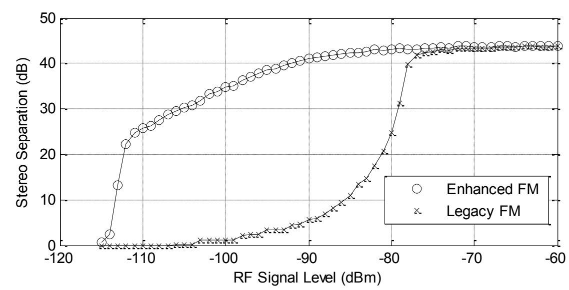

Figure 6 Stereo separation versus RF signal level

The first FME architecture has been implemented with low complexity on a commercial FM receiver chip. With the FFT and IFFT implemented in a hardware accelerator, the other FME algorithms can be implemented in less than 30 MHz on an ARM Cortex M3 microcontroller. A digital signal processor should take less than 15 MHz to implement the FME algorithms including the FFT and IFFT.

8. CONCLUSION

Two FM receiver architectures and three digital signal processing algorithms for enhancing FM audio quality are proposed. Together they achieve very large performance improvements in both subjective audio quality and objective measurements of SNR and stereo separation, and such dramatic improvements are already realized on commercial wireless connectivity chips that include an FM receiver.

9. ACKNOWLEDGMENT

James Wihardja and David Garrett participated in technical discussions related to FME, and they also provided FM signal captures and hardware accelerators, respectively.

10. REFERENCES

[1] M. O. Al-Nuaimi, “Reduction of threshold noise in FM receivers operating in quasisynchronous field,” *Communications, Radar and Signal Processing, IEE Proceedings*, Vol. 132, No. 2, pp. 94-98, April 1985.

[2] A. M. Pettigrew and T. J. Moir, “Reduction of FM threshold effect by inband noise cancelling.” *Electronic Letter*s, Vol. 12, No. 12, pp. 1082-1084, June 1991.

[3] D. L. Lee, “Multipath noise reduction for mobile FM receiver,” *IEEE Transactions on Consumer Electronics*, Vol. 39, No. 4, pp. 773-777, September 1993.

[4] K. Taura, et al., “A new approach to VHF/FM broadcast receiver using digital signal processing techniques,” *IEEE Transactions on Consumer Electronics*, Vol. 46, No. 3, pp. 751-757, June 2000.

[5] V. V. Lukin, et al., “Robust DFT based on adaptive censored estimator for FM signal processing in non-Gaussian noise environment,” 9th International Symposium on Signal processing and its Applications, pp. 1-4, February 2007.

[6] J.-H. Chen, et al., “Systems and methods for enhancing audio quality of FM receivers,” U.S. patent application 20130195163, filed January 31, 2012, published August 1, 2013.

[7] P. Vary and R. Martin, *Digital speech transmission: enhancement, coding and error concealment*, John Wiley & Sons, ltd., West Sussex, England, 2006.

[8] L. R. Rabiner and R. W. Schafer, *Digital processing of speech signals*, Prentice-Hall, Inc., Englewood Cliffs, new Jersey, 1978.

[9] J. Thyssen, “Noise Suppression System and Method,” US Patent and Trademark Office Publication Number 20110096942, filed October 4, 2010, published April 28, 2011.

[10] J.-H. Chen, “Packet loss concealment based on extrapolation of speech waveform,” Proceedings of IEEE International Conference on Acoustic, Speech, and Signal Processing (ICASSP), Taipei, Taiwan, pp. 4129-4132, April 2009.