H.-P. Hahn, “Method for Noise Reduction
in Receiving FM Stereo Signals”
J. Audio Eng. Soc., vol. 70, no. 3, pp. 208–238, (2022 March).
DOI: https://doi.org/10.17743/jaes.2021.0064

Method for Noise Reduction in Receiving FM Stereo Signals

HANS-PETER HAHN, *AES Member*
(hphahn@fm-stereo.com)

Ingenieurbüro für Akustik und Signalverarbeitung, Griesheim, Germany

This paper describes a digital signal processing method for reducing interference when receiving an analog FM stereo signal. The proposed method uses the left and right audio signals of a stereo receiver as input signals. Unlike conventional FM receiver strategies that reduce the stereo separation broadband or in frequency bands to keep the noise at a bearable level, here the noise is reduced to a quality comparable to mono while at the same time preserving the stereo separation and frequency response. This is achieved by applying signal processing rules derived from the observation of matrixed source signals recorded in intensity, time-of-arrival, and equivalence stereophony. The main part of the noise reduction is based on lowering the magnitude spectrum of the disturbed difference signal to the level of the sum signal, eliminating the excessive width of the stereo base caused by noise. The signal processing method is compatible with the FM stereo transmission standard and applicable worldwide.

0 INTRODUCTION

FM stereo transmissions are done worldwide according to the recommendation of the International Telecommunication Union Recommendation (ITU-R) BS.450 [1] in sum/difference matrixed form instead of left/right to be downward compatible for FM mono receivers. Reception of an FM stereo signal with its left (L) and right (R) channels is much more susceptible to noise and interference than mono reception of the same signal. While mono reception of a stereo signal only needs the sum signal (L+R) with audio frequencies up to 15 kHz, stereo decoding requires dematrixing the sum and difference signals to get L and R. The difference signal (L–R) ranging from 23 to 53 kHz carries more noise and interference than the sum signal because of the FM transmission method. Common receiver strategy for noise reduction (NR) is to reduce the level of the difference signal in poor reception conditions to obtain a tolerable audio signal-to-noise ratio (SNR) at the expense of stereo separation. Here the reduction of the difference signal is carried out in broadband or frequency ranges. Additionally, higher audio frequencies in the channels L and R are often attenuated to reduce noise (see Sec. 6).

This paper describes a digital signal processing method for NR [2] derived from the matrixing/dematrixing rules defined in [1]. The proposed method reduces noise to a quality comparable to mono while at the same time preserving the stereo separation and frequency response.

The presented method uses only the left and right audio signals as inputs for processing. This means that audio signals can be processed after reception or even after recording. To fully exploit the benefits of the method, receiver functions such as stereo-blend and High-Cut should be disabled. For a receiver using this technology, the stereo coverage area of a transmitter is adjusted to the mono coverage area. The paper gives deeper insight into the signal processing steps to reduce interference and leave undisturbed signals audibly unchanged.

In the first step, the performance of current FM stereo transmission technology is presented. Next, typical stereo recording techniques and their impact on the matrixed stereo signals are discussed; from this, processing rules for NR in the time and frequency domains are derived. An overview of signal processing in a block diagram is given. Finally the NR capability of the proposed method is addressed and compared to other methods. The types of interference that can be reduced are listed. A link to audio examples is provided.

1 CURRENT FM STEREO TRANSMISSION TECHNOLOGY

The pilot-tone system described in ITU-R BS.450 [1] is used to transmit stereo signals from the FM transmitter. At the transmitter site, the left and right audio signals are stereo encoded. Fig. 1 shows a simplified signal processing of a

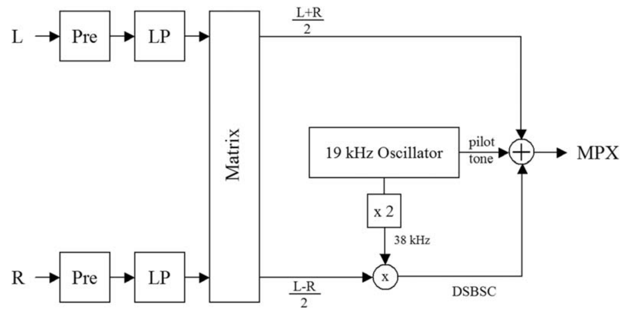

Fig. 1. Stereo encoder with pre-emphasis (Pre) and low-pass filter (LP).

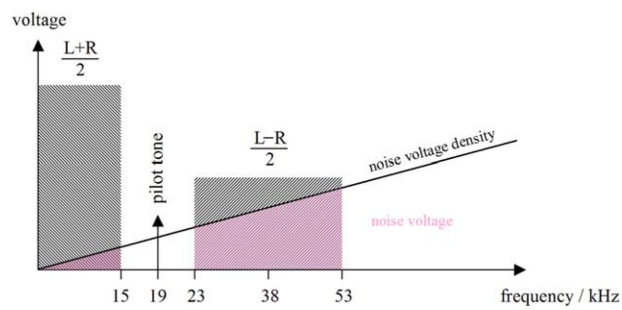

Fig. 2. MPX-spectrum: noise voltage density and noise voltage (pink area), sum and difference signals (shaded) [2].

stereo encoder. The audio signals L and R first subjected to a pre-emphasis and then are matrixed according to the matrixing rules:

$$sum = \frac{L+R}{2} \text{ and } difference = \frac{L-R}{2}. \quad (1)$$

The resulting sum signal (L+R)/2 is added to the pilot tone signal. The difference signal (L–R)/2 after matrixing modulates a frequency of 38 kHz, which is in phase with the pilot tone. The modulation is double-sideband suppressed carrier (DSBSC). The DSBSC signal is added to the sum signal and pilot tone to give the frequency multiplex (MPX) signal.

The MPX signal and, if necessary, additional signals (not shown in Fig. 1), such as Radio Data System (RDS)/Radio Broadcast Data System (RBDS), modulate an FM transmitter's high-frequency carrier signal in its frequency. The RF signal is transmitted via antenna.

At the receiver site, the RF signal from the receiving antenna is FM demodulated to get the MPX signal. The FM-demodulation process converts a frequency-constant noise power density in the RF-range into a parabolic noise power density in the MPX band [3]. Taking the root of the power, frequency-proportional noise voltage density [4] is obtained as shown in Fig. 2.

A mono receiver evaluates only the sum signal (L+R)/2 in the baseband extending up to 15 kHz. In a stereo receiver, the MPX signal is fed into a stereo decoder. The MPX stereo decoder principle shown in simplified form in Fig. 3

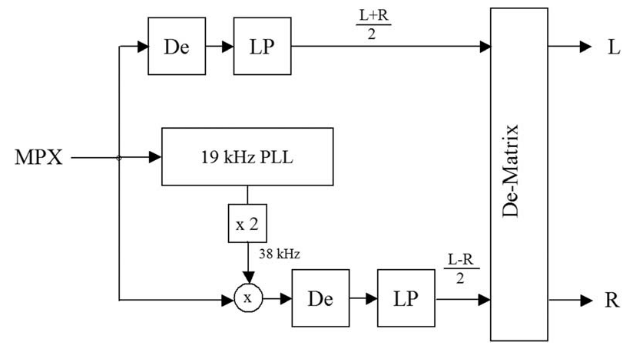

Fig. 3. Block diagram of an MPX stereo decoder with de-emphasis (De) and low-pass filter (LP).

is explained below.

In the stereo decoder, a frequency doubling of the pilot tone signal takes place and hence a recovery of the 38 kHz carrier frequency of the difference signal, happens. The stereo decoder demodulates the double-sideband modulated difference signal and thus recovers the signal (L–R)/2. The sum signal (L+R)/2 is recovered directly from the baseband. According to the dematrixing rules, the decoder recovers the pre-emphasized L and R signals again:

$$L = sum + difference \text{ and } R = sum - difference. \quad (2)$$

L and R signals are then subjected to a de-emphasis, which compensates the transmitter-side pre-emphasis. The original signals L and R are thus available.

The noise voltage density of the FM-demodulated MPX signal increases in proportion to frequency. For mono reception up to 15 kHz, the de-emphasis nearly compensates for the noise above the corner frequency of the de-emphasis network of 2.12 kHz (75µs) and 3.18 kHz (50µs), as shown in Fig. 4 (red curves).

Immediately above the FM threshold (discussed in Sec. 1.1) and far enough below the receivers maximum achievable SNR, the stereo noise is 22.2 dB above mono noise, calculated over the entire audio band up to 15 kHz with a de-emphasis of 75 µs or 20.75 dB with 50-µs de-emphasis (see Appendix A for calculation). The respective stereo noise spectrum is different from the mono noise spectrum, as can be seen from Fig. 4.

The lower and upper sidebands of the (L–R) signal have different spectra. With the DSBSC demodulation process, lower and upper sidebands' noise are added. As a result, the noise of a stereo-decoded signal differs from mono noise depending on the frequency. Stereo noise emphasizes lower audio frequencies.

Fig. 5 shows the difference of the stereo and mono noise curves in Fig. 4. This difference is called the mono-gain of SNR. Here it is shown just above the FM threshold. At 1 kHz the difference is 34.6 dB; at 15 kHz, there is a gain of 12.0 dB. The mono-gain represents the frequency-dependent NR that is obtained when the receiver is switched from stereo to mono at the expense of a total loss of stereo separation. The NR method described in this paper approximates

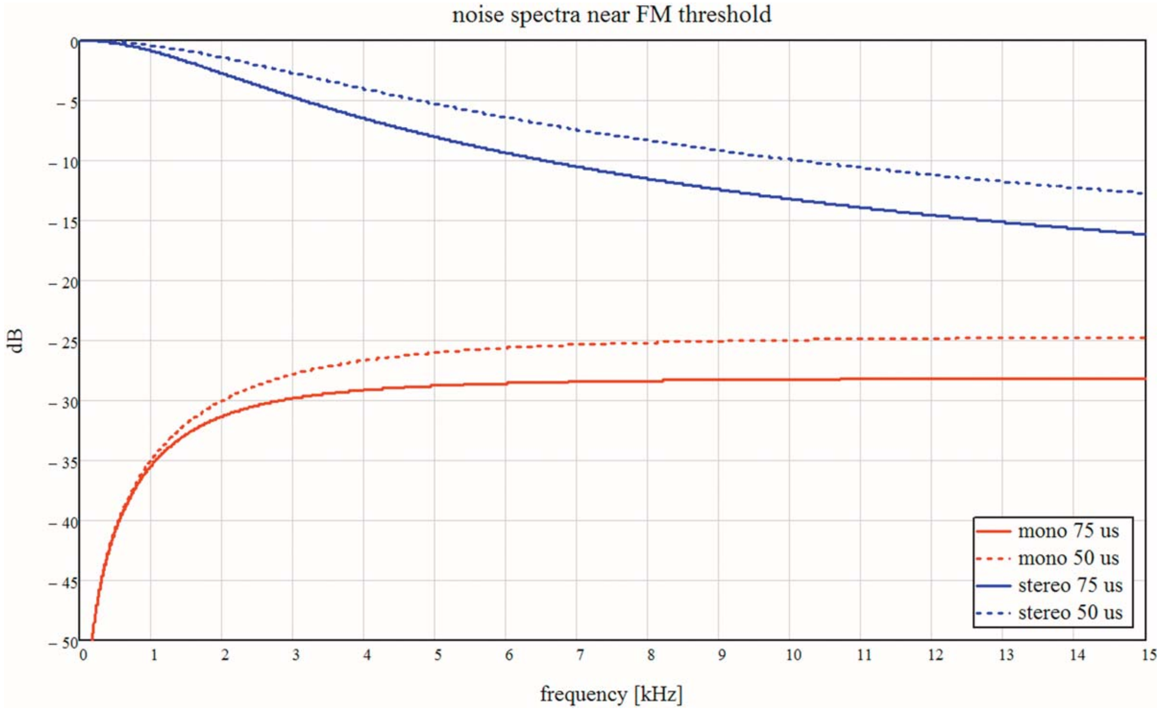

Fig. 4. Noise power spectra near the FM threshold (see Fig. 6) after decoding and de-emphasis, adjusted for 0-dB stereo noise at DC. From top to bottom: stereo 50 μs (blue dotted), stereo 75 μs (blue), mono 50 μs (red dotted), and mono 75 μs (red).

these values while maintaining stereo channel separation. In optimal reception conditions, the FM-specific noise is dominated by the inherent noise of the audio stages of the receiver. At this point, the mono-gain typically reduces to a frequency-flat 2–5 dB (depending on the receiver; e.g., Fig. 7).

The achievable SNR just above the FM threshold will now be examined.

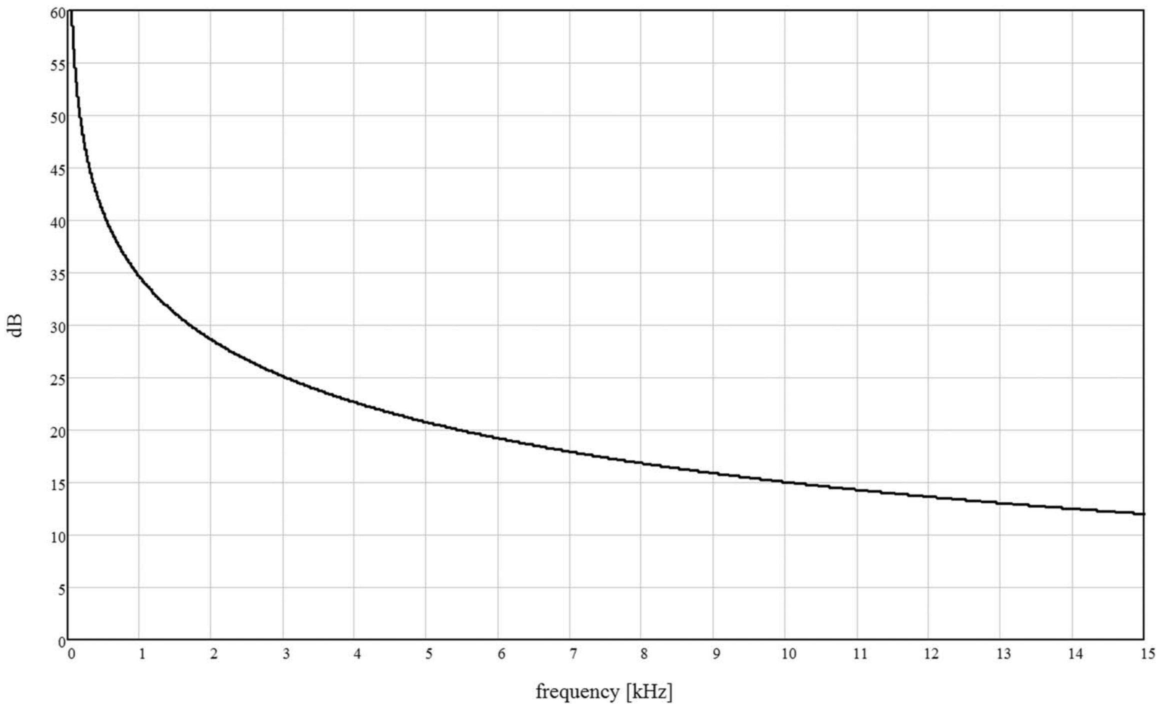

Fig. 5. Mono-gain in signal-to-noise ratio (SNR) over stereo, just above the FM threshold.

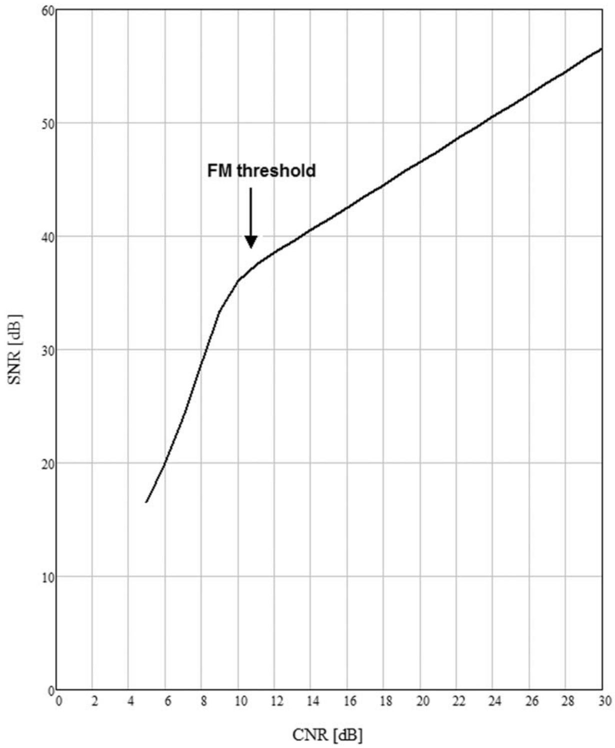

Fig. 6. Mono signal-to-noise ratio (SNR) after FM demodulation without de-emphasis as a function of carrier-to-noise ratio (CNR). For β= 5, the FM threshold level is at 11 dB CNR.

1.1 Receiver Behavior According to Current Technology

The FM pilot-tone system described in ITU-R BS.450 should first be considered in theory with respect to random FM noise. The monophonic audio SNR_{FM} prevailing after the FM demodulation with respect to +/–75 kHz frequency deviation without consideration of a pre/de-emphasis can be approximated [5] by the following formula:

$$SNR_{FM} = 3\beta^2 (\beta + 1)\, CNR, \tag{3}$$

where $CNR = \frac{A^2}{2B_T N_0}$. β is the FM modulation index (frequency deviation divided by audio bandwidth), A is the amplitude of the carrier signal, $N_0/2$ is the two-sided spectral noise power density with white noise, and B_T is the radio frequency transmission bandwidth. It can be estimated using the Carson formula by way of $B_T = 2\,(\beta{+}1)\,W$, where W is the audio signal bandwidth.

Eq. (3) applies to normal reception conditions above the FM threshold, below which the signal quality decreases rapidly [3]. Fig. 6 shows the mono SNR as a function of the radio-frequent CNR, calculated according to [6].

At a typical FM threshold of 11 dB, the mono SNR without de-emphasis is 37.5 dB [3]. Considering the NR by the de-emphasis [3] of about 13 (75 µs) or 10 dB (50 µs),

$$SNR_{FMmono} = 50.5dB\,(75\mu s)\,or\,47.5\,dB\,(50\mu s). \tag{4}$$

In the stereo case, take the mono SNR related to a peak deviation of 67.5 kHz (the pilot tone takes 10% of the peak deviation) of 49.6 (75 µs) or 46.6 dB (50 µs) and subtract the mono-gain of 22.2 (75 µs) or 20.75 dB (50 µs) to get an approximately

$$SNR_{FMstereo} = 27.4dB\,(75\mu s)\,or\,25.8\,dB\,(50\mu s). \tag{5}$$

Fig. 7 shows the FM quieting curves of a HiFi tuner with a de-emphasis of 50 µs. For a 75 µs de-emphasis, the mono noise curve shifts down by 3 dB, and the stereo noise curve shifts down by 1.6 dB. FM threshold occurs at an RF level of about 1 µV. In some Hi-Fi receivers, but in particular in car radios; some smartphones; and FM receiver chips [7–9], the level of the difference signal is reduced automatically or manually (stereo-blending) to keep the stereo noise at a bearable level (see the example below of the dotted green curve), at the expense of stereo separation (a reduction of the difference signal causes an approximation of L to R via the dematrixing rule). The level reduction of the difference signal can be broadband or in frequency ranges, e.g., in the high frequencies, and depends on the extent of external signals, external criteria, or an estimate of the interference signal. In comparison to this, the NR method described here reduces the stereo noise (green) to mono level (red curve) over the entire audio frequency band while maintaining stereo channel separation. For a discussion of classical and modern processing methods in comparison with the proposed one, see SEC. 6.

In the receiver, the mono-gain in SNR [N(stereo) – N(mono)] decreases with the rising RF level, which can be seen in Fig. 7. At the FM threshold of 1 µV, the mono-gain in SNR is at its maximum. At high RF levels, the inherent noise of the receiver, including the audio circuits after stereo decoding, limits the achievable SNR.

Further actions of receivers to reduce the audibility of interferences in the audio frequency range or MPX range include lowering of the higher audio frequencies (High-Blend, High-Cut) during strong noise and volume-reduction or muting (muting, noise blanker) during strong interference. These also have an effect on the sum signal (mono signal). Regarding types of interference other than random noise, worth mentioning are co-channel and adjacent channel interference, multipath interference, interference from machine noise, and interference from digital signals in the immediate frequency environment of the useful FM spectrum.

2 DERIVING THE RULES FOR THE NR

This section looks at the various types of stereophony in which a stereo signal can be recorded before it is transmitted in FM stereo. Specifically this section will focus on the matrixed stereo signals and derive rules for their behavior, from which signal processing steps are developed. The goal here is to design the signal processing rules in such a way that interference is reduced and the original signal is restored. Additionally, undisturbed signals should not be audibly altered by the processing method. This would allow the signal processing method to be applied to all kinds of stereo signals.

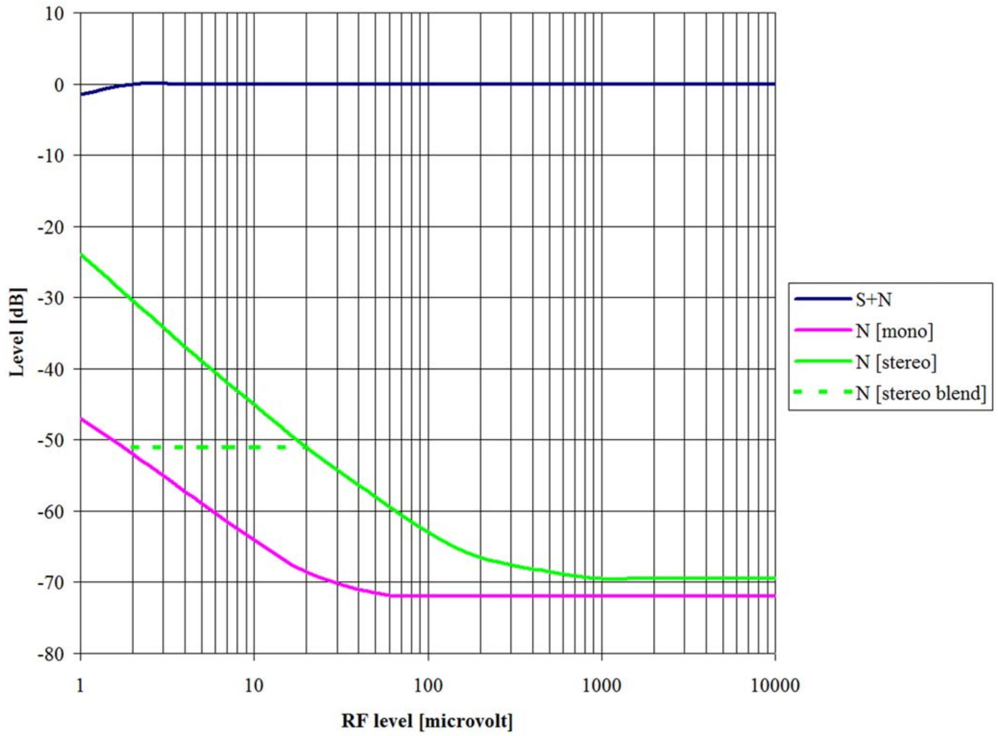

Fig. 7. FM quieting curves of a HiFi tuner with a de-emphasis of 50 µs, related to a frequency deviation of +/–75 kHz = 0 dB.

2.1 Intensity Stereophony

In the case of pure intensity stereophony (IS), a source signal (a musical instrument or a voice) is mapped to a virtual spot on the stereo base by splitting the source signal in a certain ratio between the left (L) and right (R) channels. This is called amplitude panning. The virtual source location is defined by the relative levels of the L and R loudspeaker signals. A corresponding signal splitting also takes place in intensity stereophony (IS) recordings of a sound stage with coincident microphones. Here, the level differences result from direction-dependent microphone characteristics. The signals in L and R are equal to each other in time/phase.

During reproduction, human hearing can use the level differences between the left and right loudspeaker signals to determine the direction of the auditory event. With a level difference of about 18 dB, an auditory event is located at the extremities of the stereo base [10]. The localization accuracy of IS is good, even excellent, at the positions of the speakers [11]. However IS is not able to give a clear impression of space and reverberation [12].

The stereo base extends from the far left (R = 0) to the center (L = R) up to the far right (L = 0). For FM stereo broadcasting, the audio signals L(t) and R(t) are matrixed.

A sum signal σ(t) and difference signal δ(t) are created. For L, R, σ, and δ, for the sake of simplicity, the time dependence is henceforth assumed and omitted in the remaining representation. The matrixing specification is as follows:

$$\sigma = \frac{L+R}{2} \quad and \quad \delta = \frac{L-R}{2}. \tag{6}$$

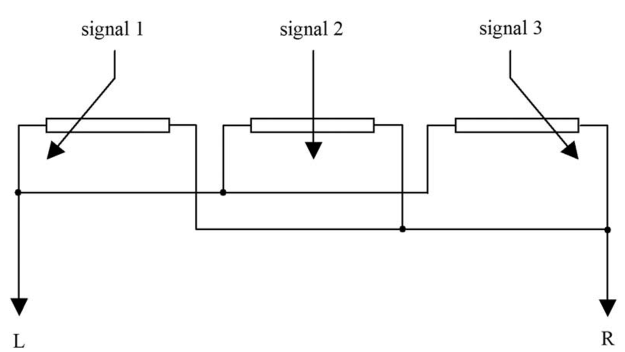

Fig. 8. Pattern of signal splitting in channels L and R with pure intensity stereophony (IS) [2].

A de-matrixing takes place on the receiving side:

$$L = \sigma + \delta \quad and \quad R = \sigma - \delta. \tag{7}$$

2.1.1 IS With a Single Source Signal

First a single signal (e.g., signal 1 in Fig. 8) should be considered.

If it is assumed that on the transmitting end there is no excess width of the stereo base, in other words R = 0 and L = 0 represent the extreme locations of the stereo base, then results for

$$R = 0: \quad \delta = \sigma, \tag{8}$$

and for

$$L = 0: \quad \delta = -\sigma, \tag{9}$$

deriving:

$$|\delta| \;=\; |\sigma|\,. \tag{10}$$

For each spot mapped within the stereo base, regarding the single signal follows

$$|\delta| \;\leq\; |\sigma|\,, \tag{11}$$

where the equality holds true for the cases R = 0 and L = 0.

Instead of comparing |δ| and |σ|, δ can be compared with the envelope curve of σ. This makes the comparison independent of the polarity of δ. This weaker condition still meets Eq. (11). In this context, |σ| can be interpreted as an envelope curve [Fig. 9(c)]. The following can be defined:

- **Rule 1 for signal behavior in the time domain**.

  $$|\delta(t|) \leq |\sigma(t)|. \tag{12}$$

  At any given time, the absolute value of the difference signal is smaller than that of the sum signal or is at most equal to it. The difference signal lies within the envelope of the sum signal. The envelope is supported by relative maxima/minima of the sum signal.

Before the frequency domain is discussed, some terms should be clarified. The signal processing described in this paper works block-wise and uses the discrete Short-Time Fourier Transform and its inverse.

Each block (e.g., 4,096 samples) is transformed into the frequency domain by the discrete Short-Time Fourier Transform. In the frequency domain, the complex frequency spectrum is composed of discrete spectral lines. Each spectral line with its associated frequency can be interpreted as a complex vector with a real and imaginary part or in polar form, with a magnitude and phase.

In the following, the terms "spectrum" and "signal spectrum" are understood as magnitude spectrum, as long as not specified in more detail. For further information, the magnitudes of spectra are considered to be logarithmic and are scaled in decibels. Now the sum and difference signals in the frequency domain are examined—for the sake of clarity, in the continuous-time form.

The real valued time domain signals σ(t) and δ(t) are transformed into the frequency domain:

$$F\left\{\sigma\left(t\right)\right\} = \Sigma\left(\omega\right) = \mathrm{Re}\left\{\Sigma\left(\omega\right)\right\} + j\mathrm{Im}\left\{\Sigma\left(\omega\right)\right\}, \tag{13}$$

or in polar form

$$F\left\{\sigma\left(t\right)\right\} = \left|\Sigma\left(\omega\right)\right| e^{j\measuredangle\Sigma(\omega)}, \tag{14}$$

with the magnitude spectrum of σ(t)

$$\left|\Sigma\left(\omega\right)\right| = \sqrt{\mathrm{Re}^2\left\{\Sigma\left(\omega\right)\right\} + j\mathrm{Im}^2\left\{\Sigma\left(\omega\right)\right\}}, \tag{15}$$

and phase spectrum

$$\measuredangle\Sigma\left(\omega\right) = \arctan\left(\frac{\mathrm{Im}\left\{\Sigma\left(\omega\right)\right\}}{\mathrm{Re}\left\{\Sigma\left(\omega\right)\right\}}\right), \tag{16}$$

according

$$F\left\{\delta\left(t\right)\right\} = \left|\Delta\left(\omega\right)\right| e^{j\measuredangle\Delta(\omega)}. \tag{17}$$

If the difference signal δ(t) has opposite polarity, as shown in Fig. 9(b) compared with Fig. 9(a),

$$\begin{aligned} F\left\{-\delta\left(t\right)\right\} &= -F\left\{\delta\left(t\right)\right\} = \left|\Delta\left(\omega\right)\right| e^{j\measuredangle\Delta(\omega)} e^{j\pi} \\ &= \left|\Delta\left(\omega\right)\right| e^{j\measuredangle\Delta(\omega)+j\pi}, \end{aligned} \tag{18}$$

with $e^{j\pi} = -1$.

A negative sign of a time domain signal is expressed in the phase spectrum by an additional phase angle of π. Thus multiplying the difference signal spectrum with exp(jπ) or exp(−jπ) would flip L and R. The magnitude spectrum is always positive.

In the frequency domain, information about the location of an auditory event within the stereo base is contained in the phase and magnitude spectra of the sum and difference signals. While in the time domain, a different sign of the difference signal compared to the sum signal assigns the sound source to the right half plane of the stereo base; in the frequency domain, this sign is part of the phase spectra of the difference and sum signals. Within a half plane, the relation of the sum and difference magnitude spectra levels gives detailed information about the location of the sound source. The following is defined:

- **Rule 1 for signal behavior in the frequency domain**.
  Magnitude spectra:

  $$|\Delta(\omega)| \leq |S(\omega)|, \tag{19}$$

  Power spectra:

  $$|\Delta(\omega)|^2 \leq |S(\omega)|^2. \tag{20}$$

  At each frequency, the magnitude/power of the difference signal is smaller than the magnitude/power of the sum signal or is at most equal to it.

so that:

- **Rule 1 for signal processing in the frequency domain**.
  The magnitude of each spectral line of the disturbed difference signal spectrum is reduced to the corresponding magnitude of the sum signal spectrum. The phase spectrum is further processed unchanged.

The reduction of the difference signal magnitude spectrum to the sum signal magnitude spectrum at some frequencies corresponds to a return of the signal from outside the stereo base back inside the stereo base. At the same time, this reduces the noise. It should be noted that other spectral lines may already be at or below the value of the sum signal spectrum and are not changed. The application

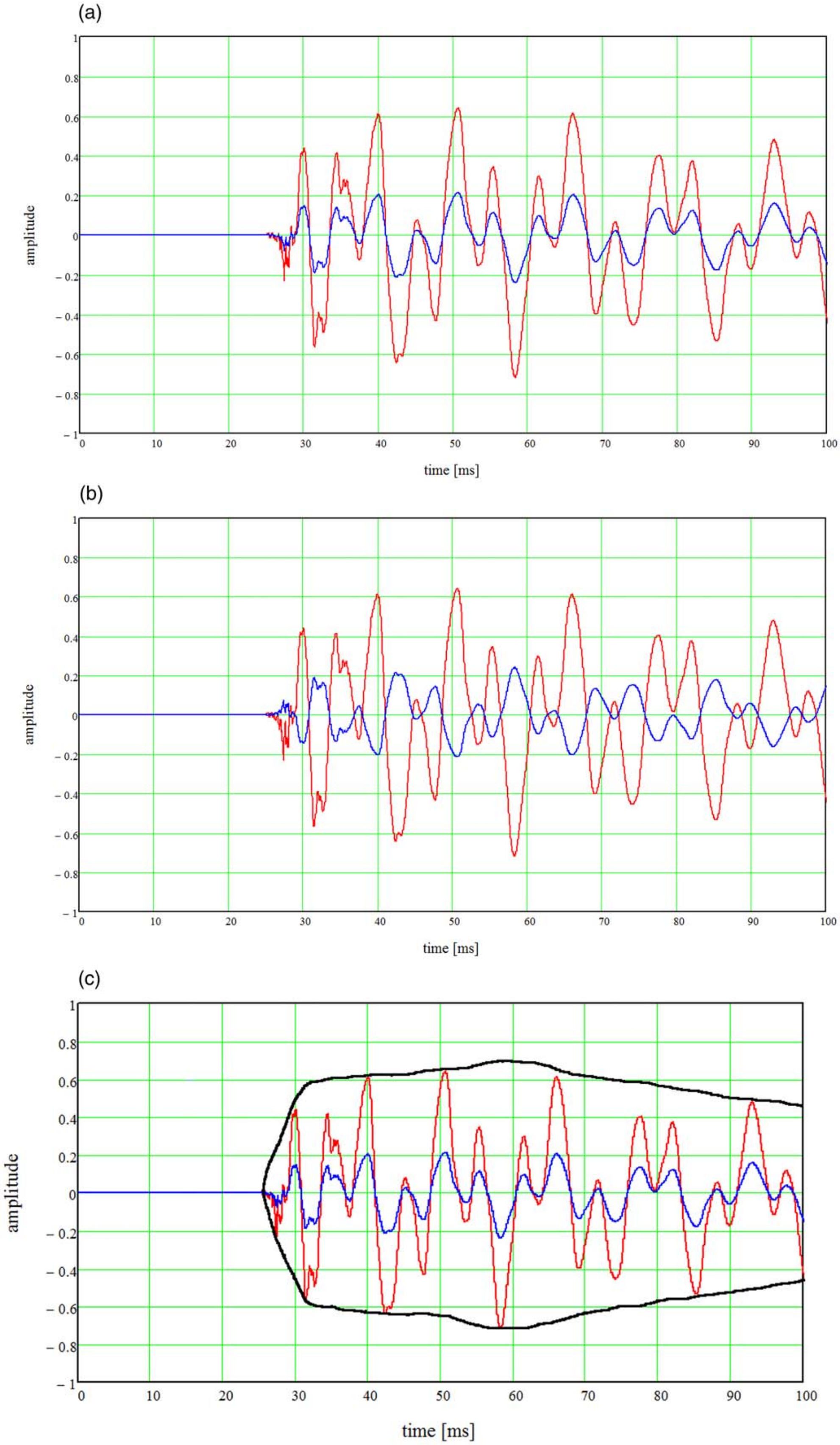

Fig. 9. (a) Example of a sum signal (red) and difference signal (blue) at pure intensity stereophony and a predominant distribution of a signal into the left channel L. (b) Example of a sum signal (red) and difference signal (blue) at pure intensity stereophony and a predominant distribution of a signal into the right channel R. The difference signal has an opposite polarity compared to (a). (c) The difference signal δ (blue) always lies within the envelope of the sum signal σ (red), independent of its polarity. The envelope (black) is supported by relative maxima/minima of the sum signal.

of Rule 1 in the frequency domain represents the largest share of NR.

2.1.2 IS With Multiple Source Signals

Back to the time domain, multiple signals (e.g., signal 1 + signal 2 in Fig. 8) are now considered, instead of a single signal. Here the situation is somewhat different.

From Figs. 10 and 11, it can be seen that the difference signal can have a higher amplitude than the sum signal or its envelope. If, as here, the individual signals are localized at the extremities of the stereo base (R = 0 and L = 0), the difference signal level is highest. Furthermore the signal constellation with a frequency ratio of 3:1 represents a dominant case that leads to significant amplitude differences of the sum and difference signals. This frequency ratio often occurs as the ratio of a fundamental to its third harmonic. In most cases of complex signal constellations, the ratio of the difference signal to the sum signal is no higher than 1.4.

It is also apparent that the maximum of the sum and difference signal does not necessarily have to be concurrent. With a frequency ratio of 3:1, the time difference between the extreme values of the sum signal and difference signal corresponds to a half period of the higher frequency signal. To estimate an upper limit for the time difference of these extreme values, assume a frequency ratio of 50:150 Hz, although the contribution of low frequencies to the difference signal is small (the low frequency range is usually transmitted in mono). A range of +/–3.5 ms therefore covers almost all such effects.

In intensity stereophonic signal constellations, for all mapped spots within the stereo base applies:

- **Modified Rule 1 for signal behavior in the time domain**.

  $$|\delta(t)| \leq k_{IS}|\sigma(t)| \qquad \text{within a time window of } \tau_{IS}, \tag{21}$$

  with k_{IS} = amplitude factor for multiple signals, e.g., 1.4; and with τ_{IS} = peak hold time, e.g., +/–3.5 ms.

  The difference signal is within an envelope curve. The envelope is based on the relative maxima/minima of the sum signal multiplied by a factor of k_{IS}. Each newly detected extreme value—multiplied by k_{IS}—is kept for a peak hold time of τ_{IS}. Since an extreme value can occur first in the sum signal or the difference signal, the hold time also applies to periods before the observation time.

For intensity stereophony, this results in:

- **Modified Rule 1 for signal processing in the time domain**.

  For each block the disturbed difference signal is clipped to the envelope of the sum signal, wherein the envelope takes into account signal shifts and amplitude increases. To compute the envelope, the relative maxima (momentary peak values) of the absolute value of the sum signal are held (peak hold) for a time τ_{IS}. This hold time is also applied to the time before a peak value occurs. The resulting envelope curve is increased by the factor k_{IS} (e.g., 1.4) and applied symmetrically around the zero line (DC). It is necessary that the block length (in this case about 100 ms) covers the time difference between the extreme values of the sum and difference signals.

Especially at temporarily very low SNRs, clipping the difference signal reduces the predominant noise at the expense of useful signal components, which have been partially cut off. Sec. 2.4 discusses clipping in more detail and shows that it can be used to good effect in certain situations.

2.1.2.1 Multiple source signals and processing in the frequency domain. This section discusses whether frequency domain processing, as defined above, can also be applied to multiple source signals. For this, the L and R channel signals, which result after panning (see Fig. 8), are first considered. Then it is checked whether there are cases where the difference signal can exceed the sum signal (after matrixing).

By panning multiple source signals to the middle range of the stereo base, the difference signal is kept low compared to the sum signal, and Rule 1 in the frequency domain is valid.

The unfavorable case where at least two source signals are panned into the opposite extremities of the stereo base is considered. Here Rule 1 and the frequency domain processing apply if the spectra of the various source signals (see Fig. 8) have no common spectral lines, as in Fig. 10.

In the case of multiple source signals with common spectral lines (frequencies), a single frequency, representative of every line in the entire spectrum, can be examined. For every spectral line (frequency) with the same phase/polarity in channels L and R, the matrixing rules result in a difference signal magnitude smaller than the sum signal magnitude. This extends to the case of a phase difference between $-\pi/2$ and $+\pi/2$. Such a phase difference causes the resulting difference vector to have a smaller magnitude than the sum vector.

There is a conceivable case in which Rule 1 could be violated. If source signals with common spectral lines (frequencies) are panned to opposite positions of the outer stereo base, and a common spectral line occurs in channels L and R with a phase difference between $\pi/2$ and $3\pi/2$, this may cause the difference signal spectral line vector to have a larger magnitude than the sum vector. Since this is a rare case and of rather theoretical consideration, signal processing in the frequency domain, as defined earlier, can also be performed for multiple source signals.

Signal processing will be explained using three examples:

In the outer area of Fig. 12, the noise in the difference signal (blue) exceeds the extended envelope (black) of the sum signal (red). A truncation of the interfering signal sections is possible. The envelope takes into account possible time shifts and amplitude increases of the signal. As you

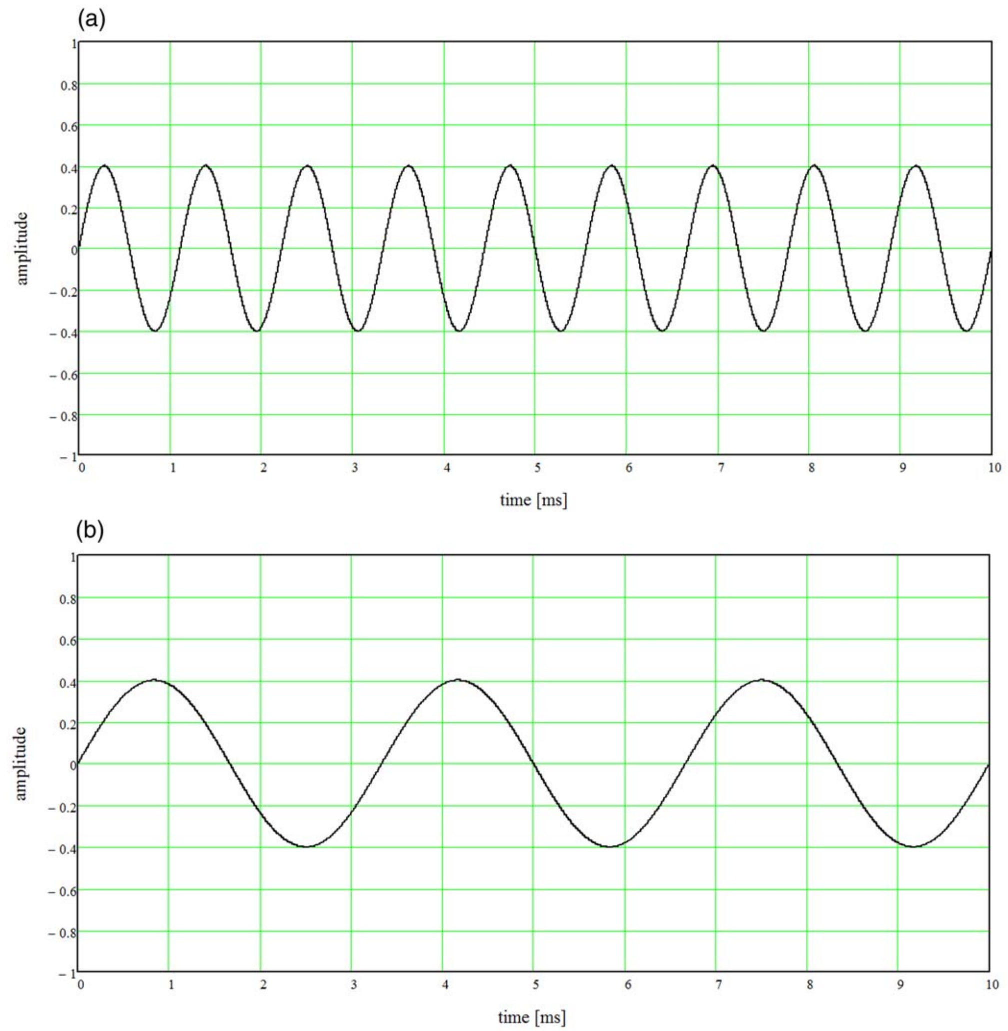

Fig. 10. (a) L = sine wave 900 Hz [2]. (b) R = sine wave 300 Hz [2].

can see, the envelope extends to both sides of the signal maximum.

Fig. 13 shows that the difference signal (blue) lies within the extended envelope (black) of the sum signal (red) A level reduction in the time domain is not possible.

Fig. 14 shows that an NR can be performed in the frequency domain. At some points, the disturbed difference signal spectrum (blue) can be reduced to the sum signal spectrum (red). The extent of the reduction is indicated by delta (black).

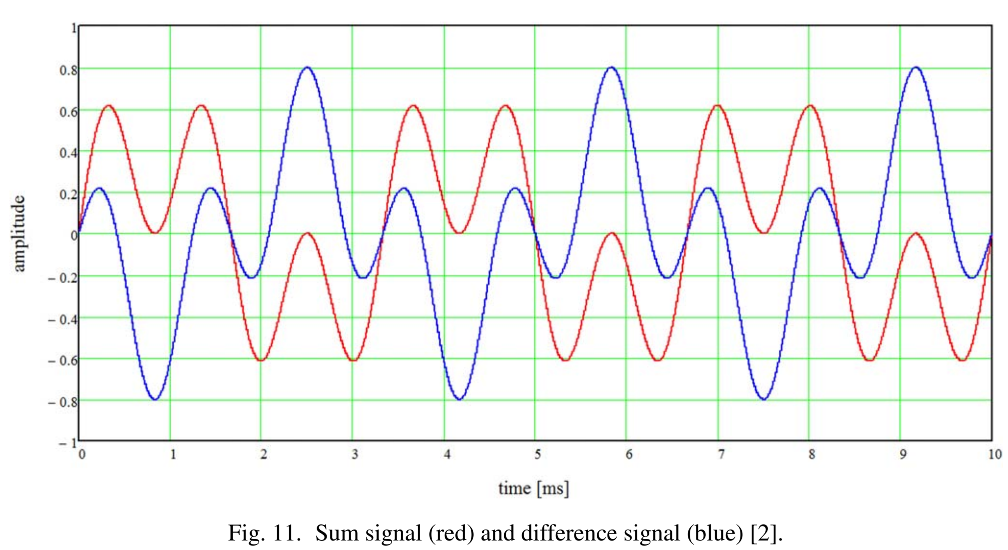

Fig. 11. Sum signal (red) and difference signal (blue) [2].

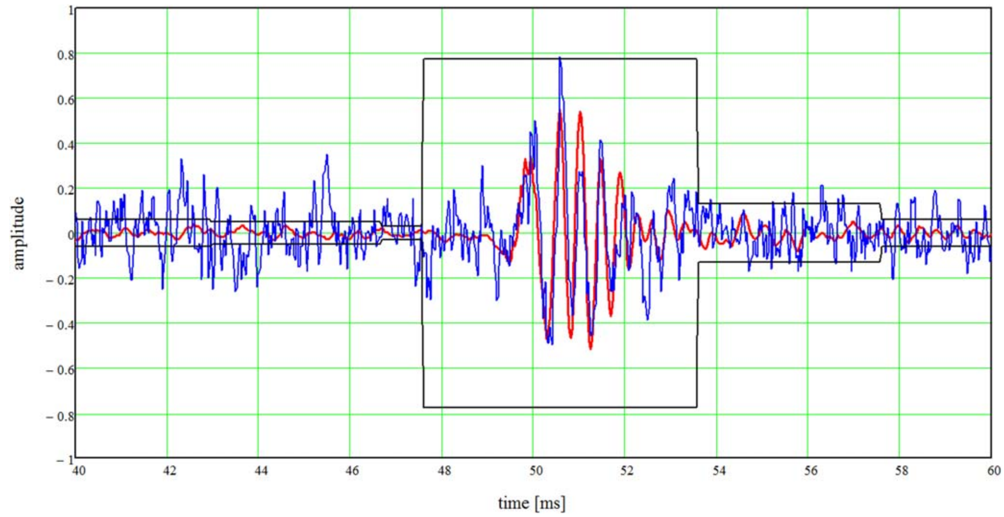

Fig. 12. Example in which noise reduction in time domain is possible.

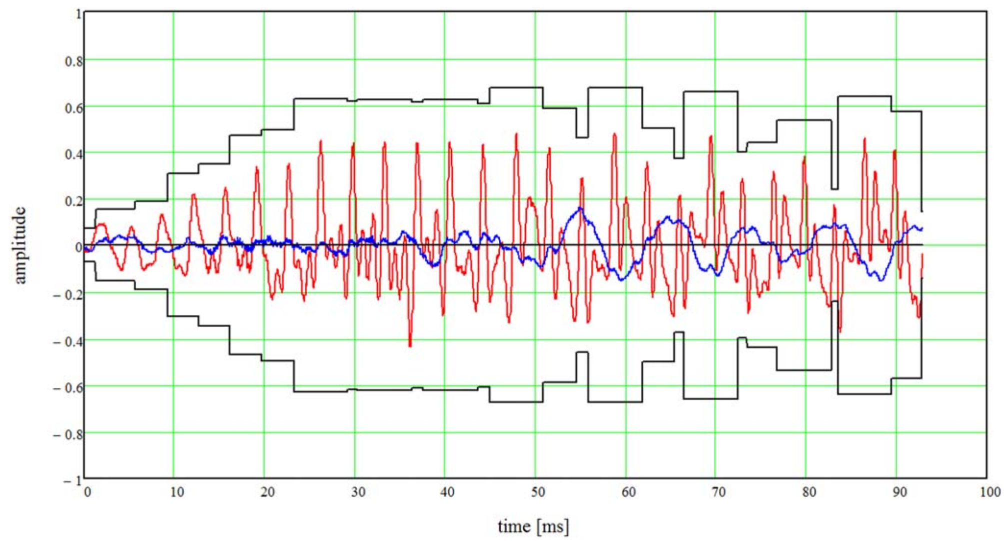

Fig. 13. Example in which noise reduction in the time domain is not possible.

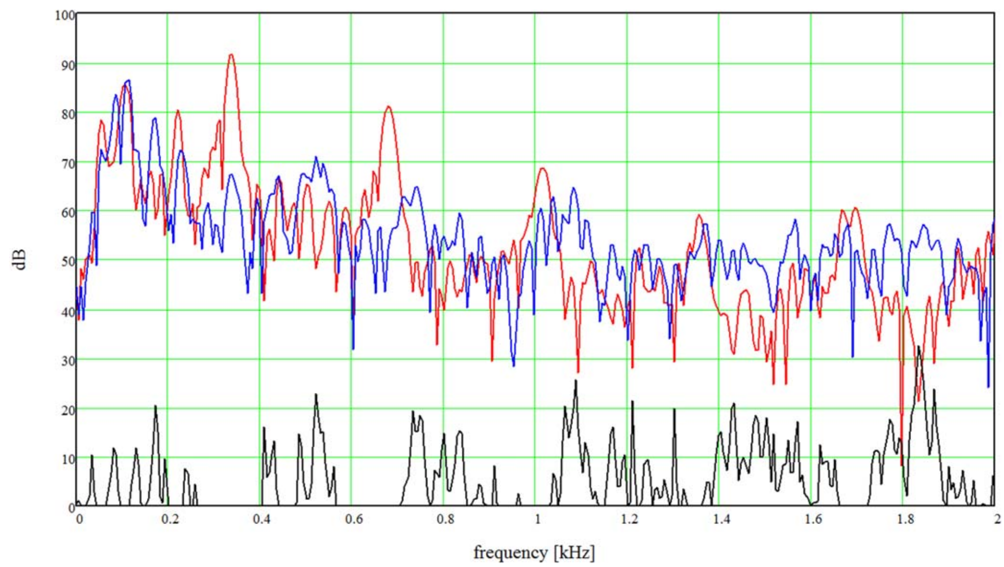

Fig. 14. Example in which noise reduction in the frequency domain is possible.

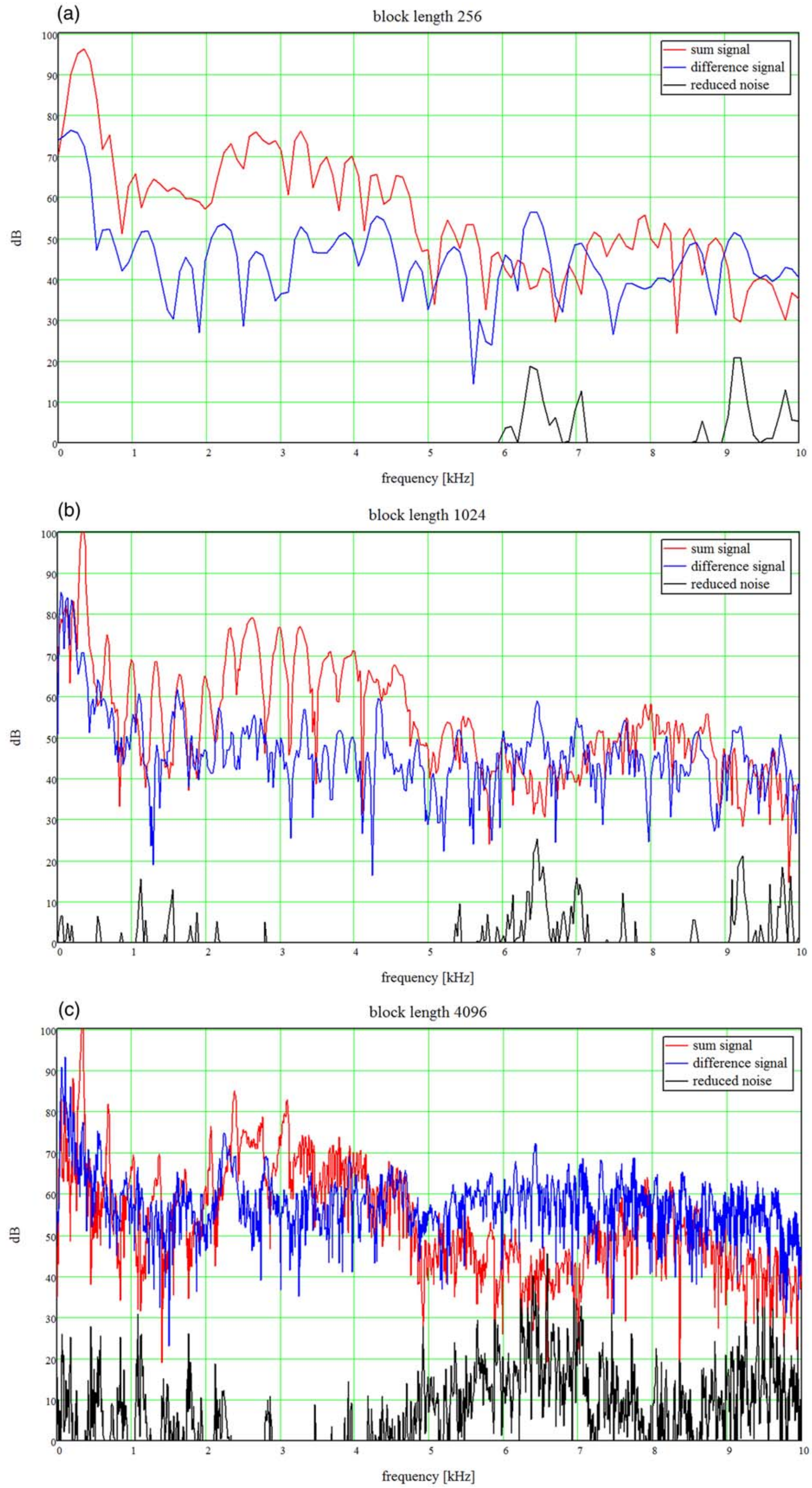

Fig. 15. (a) Block length 256. (b) Block length 1,024. (c) Block length 4,096.

With increasing frequency resolution (corresponding to increasing block length), more details of the spectra are opened, also in the form of gaps and peaks. The method exploits this situation and lowers interfering signals in the difference signal spectrum at the identified points. Fig. 15 shows signal spectra for different block lengths. The black curve is the amount of NR. As can be seen, NR in the low frequencies is best with a block size of 4,096.

However a further increasing block length leads to an adjustment of the spectra of sum and difference signals (depending on the signal characteristics) and limits the interference reduction. On the other hand, a minimum number of samples (minimum block length) is needed for a sufficient frequency resolution, even at low frequency bands. As a compromise, a block length of 4,096 at a sample frequency of 44.1 kHz was determined empirically.

Figs. 15(a)–15(c) signal spectra and block length: sum (red), difference (blue), and noise (black).

At this point, the signal processing of a complex spectral line vector shall be considered as an example. Fig. 16 shows a vector diagram with the sum and difference signal line vectors.

The stereo base span at a certain frequency is defined by the length (magnitude) of the sum signal vector. In this example, the transmitted difference signal vector length is shorter. This means this spectral line is located within the stereo base near its outer limit. FM reception interference adds to the difference signal. The resulting vector lies outside the stereo base.

Signal processing according to the proposed method reduces the resulting vector length (magnitude) back to the stereo base. In this case, a phase difference β is left over between the processed vector and original difference signal vector.

Note that every spectral line is processed individually and that the perception of sounds consisting of a bundle of spectral lines occurs within critical bandwidths of human hearing. Due to the high frequency resolution of the method, there are several spectral lines within each frequency group. The reduction of spectral lines according to the processing rule leads to an improvement of the SNR within this frequency group.

Any noise remaining after dematrixing must be seen within the critical bandwidth and may go unnoticed by human hearing because of the common perception of spectral lines within a group and the shape or spectral content of the signal.

Nevertheless, at very low SNR, remaining phase errors can lead to blurred localization of sound sources within the stereo base. Since human hearing is not very sensitive to localization [13], blurring artifacts are not recognized in most cases.

The audibility of a remaining interfering signal after NR depends on the extent to which it can be masked in the same and other channel after dematrixing. Masking takes place in the time, frequency, and spatial domains, or any combination thereof. It is dependent on the properties of the stereo signal in these three domains. Masking in the spatial domain refers to the masking of a signal of a sound source at one location within the stereo base by the signal of another sound source within the stereo base. Masking in the frequency domain is dependent on the shape of the frequency spectrum of the left and right audio signals. See Sec. 5 for more information about the audibility of artifacts at very low SNR. The complex issue of masking residual interference after dematrixing—and thus the subjective quality of the processed signal—remains a subject of further investigation.

2.2 Time-of-Arrival Stereophony and Equivalence Stereophony

With pure time-of-arrival stereophony (TAS), a sound source is recorded with displaced omnidirectional microphones. The sound travels different distances to the microphones, depending on the input direction. Within the microphone signals L and R, signals are formed that have a direction-dependent time delay Δt and the same level. In loudspeaker playback, the human auditory system can use the time difference between the left and right loudspeaker signals to determine the direction of the auditory event. With an inter-channel time delay Δt of +/–1.2 ms, an auditory event is located at the extreme points of the stereo base [10]. The localization accuracy of TAS is best in the center of the stereo base and gets worse toward the positions of the loudspeakers [11].

Recordings with TAS are made, for example, with AB technology. The captured time differences between the L and R signals add a certain airiness and ambience to the stereo reproduction.

In order to control localization only on the basis of time delay differences of +/–1.2 ms, a microphone base of at least 0.5 m is required. In the frequency domain of the sum (mono) signal, nulls can already occur at frequencies of 400 Hz at an angle of incidence of 55°. Changes in the timbre in a mono reproduction may occur. TAS main microphone setups are therefore less suited for mono playback [12].

In practice, an intensity stereophony with time-based stereophonic fractions is often used. This is also referred to as equivalence stereophony (ES). The microphone base is kept shorter than with TAS. The resulting time delay between the channels is not sufficient to cover the full stereo base during playback. This can be achieved by the additional usage of level differences. The microphones have a directional pattern and are angled outward. Localization at the extremities of the stereo base can be achieved, for instance, with a combination of a typical level difference of 6 dB and time difference of 0.5 ms [12]. For the same localization of an auditory event, the difference signal and its magnitude spectrum may be smaller than with pure IS, but for a sharper image of a sound source, a higher IS portion is preferable. This can be done, for example, with a near-coincident microphone arrangement like ORTF [14] with a microphone distance of 17 cm. ES combines the localization accuracy of IS with the ability of TAS to produce an impression of the recording space and ambience.

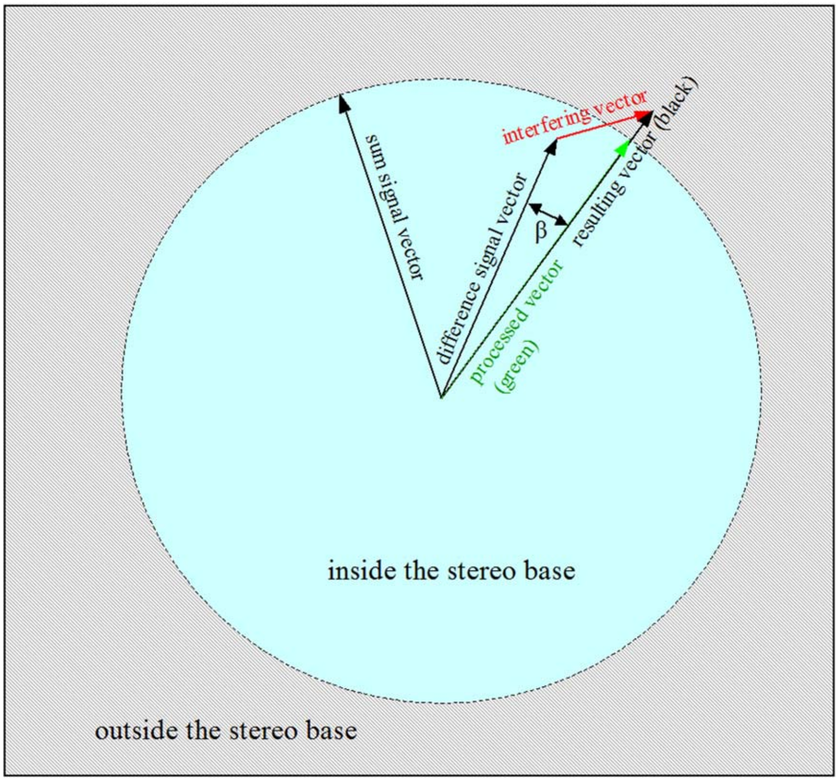

Fig. 16. Vector diagram.

Microphone setups consisting of ES main and support microphones are also classified as ES. This should also include arrangements with three microphones (e.g., Decca Tree [15], where one microphone is mixed to the center).

In the frequency domain, comb filter effects appear in the sum (mono) channel that are less pronounced than in pure TAS. Therefore ES is better suited for mono than TAS. The perception of comb filter effects is discussed in [16] and [17].

The temporal information is hidden in the phase spectra. This includes information for localization of an auditory event within the stereo base and about the temporal position of an auditory event within the block, as well as a spatial impression. The statements in SEC. 2.1.1 about the importance of the magnitude and phase spectra in stereo localization are also valid here.

The following focuses on ES, keeping in mind that ES includes portions of IS and TAS.

With ES microphone setups, the sound from a source arrives at the individual microphones with individual delay times and levels, depending on the sound input direction. The individual microphone signals are processed to a left and right audio signal according to certain aspects. Depending on the sound input direction and arrangement of the microphones, the following effects can be observed after matrixing:

- The time differences of the microphone signals manifest themselves in non-simultaneous amplitude peaks in the time domain of the sum and difference signals.
- In the time domain, the individual microphone signals are superimposed with their different time delays, generating a statistical deviation of the amplitude of the difference signal compared with the sum signal. The effect is similar to that shown in Fig. 11. This is especially true for frequencies above the bass range, in which the individual time delay results in ambiguousness of the phase (1m $\asymp$ 3ms $\asymp$ 360° at 333 Hz!).
- In the frequency domain, the matrixing of the microphone signals with their individual time delays can lead to constructive or destructive interference. This manifests as cancellation or exaggerated effects (comb filter effects) in the magnitude spectra of the sum and difference signals.

However, when recording, attention is paid for mono compatibility. That means clearly audible cancellation effects in the sum signal are avoided by choosing an appropriate microphone arrangement or carefully adjusting the level differences to the time delays of multi-microphone signals. Therefore the intensity stereophonic portion in the signal predominates, and the statistical amplitude distortion of the difference signal is limited. The following can be defined:

- **Rule 2 for Signal Behavior in the Time Domain**

$$|\delta(t)| \leq k_{ES}|\sigma(t)|, \quad \text{within a time window of } \tau_{ES} \tag{22}$$

with k_{ES} = amplitude factor, e.g., 1.4; with τ_{ES} = peak hold time, e.g., +/–3 ms.

The difference signal is within an envelope curve. The envelope is based on the relative maxima/minima of the sum signal multiplied by a factor of k_{ES}. Each newly detected extreme value—multiplied by k_{ES}—is held for a peak hold time τ_{ES}. The hold time also applies to periods before the observation time.

For ES, this results in:

- **Rule 2 for Signal Processing in the Time Domain** For each block, the disturbed difference signal is clipped to the envelope of the sum signal (especially with transients), wherein the envelope takes into account signal shifts and amplitude increases. To compute the envelope, relative maxima (momentary peak values) of the absolute value of the sum signal are held (peak hold) for a time of τ_{ES}. This hold time is also applied to the period before a peak value occurs. The resulting envelope curve is increased by the factor k_{ES} (e.g., 1.4) and applied symmetrically around the zero line (DC).

Signal processing in the time domain according to Rules 1 and 2 is similar. Peak hold times are about +/–3 ms, which accounts for microphone distances of 1 m, as with the Decca Tree.

As mentioned in connection with Rule 1, Sec. 2.4 discusses clipping in more detail and shows that it can be used to good effect in certain situations.

2.2.1 Signal Processing in the Frequency Domain

In ES, an off-center sound event is represented in the sum and difference signals as a superposition of two time-shifted signals. For the observation in the frequency domain, it is important that all time-shifted components can be transformed into the frequency domain in one step. This is the case if the block length (e.g., 100 ms) covers the major time-delay differences, i.e., time-shifted signal components occur in the same block.

Signal processing in the frequency domain in compliance with Rule 1 can also be performed for ES, except for constructive or destructive interference effects, which are discussed in Sec. 2.3. The difference signal phase spectrum is further processed unchanged. Looking at the frequency domain, the previous signal processing steps at IS and ES can be summarized with regard to stereo separation.

In the frequency domain, information about the location of an auditory event within the stereo base is contained in the phase and magnitude spectra of the sum and difference signals. While in the time domain, a different sign of the difference signal compared to the sum signal assigns the sound source to the right half plane of the stereo base; in the frequency domain, this sign is part of the phase spectra of the difference and sum signals. Within a half plane, the relation of the sum and difference magnitude spectra levels gives information about the location of the sound source.

By reducing the difference signal magnitude spectrum to the sum signal magnitude spectrum, thereby reducing noise, the spectral lines that emerge from the stereo base because of interference are returned to inside the stereo base. This corrects the ratio of the sum and difference signal magnitude spectra to permissible values. The magnitude spectra thus continue to provide information about the position of a sound source within a half plane of the stereo base.

The phase spectra contain information about the assignment of signal components to the left and right half planes of the stereo base. In the case of ES, they additionally provide temporal information for the localization of an auditory event within the stereo base.

The phase spectrum of the difference signal is not changed during signal processing because a phase change of a useful spectral line vector due to noise is usually negligible and small even at high levels of interference. This is because the SNR at full modulation is at least 27.4 dB, i.e., the useful audio signal power is at least 550 times the total noise power, both calculated over the entire audio frequency band (see Sec. 1.1). If it is further assumed that the noise is distributed over the entire audio frequency band but the useful signal is concentrated on individual spectral lines, the local SNR for a spectral line is even larger.

All these points, and the fact that the sum signal is not changed at any time, cause the stereo channel separation to be preserved.

2.3 Constructive and Destructive Interference in ES

Constructive and Destructive Interference effects in ES are normal but may cause unwanted changes in even the undisturbed audio signal during signal processing according to Rule 1 in the frequency domain. This section deals with, on one hand, the identification of these effects and how to maintain the quality of the undisturbed audio signal and, on the other hand, the reduction of noise and interference. Corresponding signal processing rules are defined that represent deviations from Rule 1 in the frequency domain.

2.3.1 Cancellations

Because of different signal path lengths (see Fig. 17) to the displaced microphones L and R, a cancellation happens, e.g., in the sum signal $\sigma = (L+R)/2$ if a frequency at microphone R undergoes a phase rotation of 180° with respect to microphone L and the levels at both microphones are equal. In the difference signal $\delta = (L–R)/2$, a cancellation happens if a frequency of the microphone R undergoes a delay-dependent phase rotation of 0°. This is called destructive interference.

With signals recorded in ES, frequency-selective cancellations (CANs) can occur in the sum and difference signal spectra, usually at different frequencies. CANs are not as deep as with TAS because the directivity of the microphones and their angling outward complicate the equal level criterion. In many cases, CANs in a signal spectrum are covered by diffuse components of the sound or the spectra of other

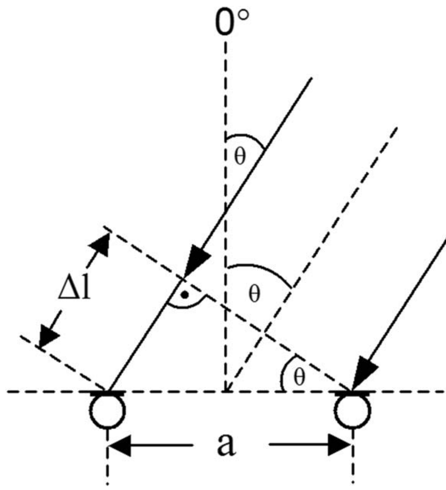

Fig. 17. AB microphone setup with time-of-arrival stereophony (TAS) [2]. Path length difference $\triangle l = a * \sin \theta$, and time difference $\triangle t = \Delta l/c$, with $c = 343$ m/s and microphone distance (a).

sources. Covering also occurs when signals of sound events do not overlap in time but are in the same block (the magnitude spectrum is independent of the temporal position of a signal in the block).

Fig. 18 shows frequency-selective CANs in the sum signal spectrum at 2.09 and 2.83 kHz. CANs occur in both spectra at different frequencies.

With a CAN in the sum signal spectrum, the processing according to Rule 1 in the frequency domain would lead to an unwanted reduction of the difference signal spectrum at the respective frequency and might even impair the sound of the undisturbed audio signal. Processing should therefore not be carried out. On the other hand, since the difference signal spectrum at this frequency may have a high interference level, a level reduction is appropriate. In order to fulfill both requirements, it makes sense to reduce the difference signal spectrum at this frequency to a spectral substitute value of the sum signal spectrum that bridges the CAN dip (see Fig. 19). Thus the NR remains in effect without distorting the useful signal. The spectral substitute value can be, for example, the median value within a given frequency interval. From this follows:

- **Rule 2 for signal processing of CANs in the frequency domain.**
  If a narrow CAN in the sum signal spectrum is identified, the value of the difference signal is reduced to the median value of the sum signal spectrum.

2.3.2 Local Level Maxima

A common effect of ES recordings is the appearance of local level maxima (LM) in the difference signal spectrum of the audio signal. Oftentimes, the IS portion of the audio signal causes LM while adhering to Rule 1 in the frequency domain. However constructive interference in the difference signal after matrixing can further increase the level at these frequencies. The exaggerations in the difference signal spectrum can amount to several decibels relative to the immediate spectral environment, and Rule 1 in the frequency domain may be violated. Signal processing would then result in undesirable attenuation of spectral lines and even a change of the undisturbed audio signal. The goal is thus both to preserve the audio quality (LM) of the original signal and reduce possible interference.

Before defining a suitable signal processing, other reasons for LM of the difference signal spectrum have to be considered, such as frequency-selective interference from equipment in the home. In this case, a lowering of the LM in the difference signal spectrum would be desirable.

An interfering signal usually has a greater effect on the difference signal than on the sum signal. In ES, an interference causing an LM in the difference signal spectrum can be detected by a lack of support from the sum signal spectrum at that frequency. In this case, frequency domain processing in compliance with Rule 1 should be carried out.

If an LM is detected in both the sum and difference signal spectra at the same frequency range, the useful signal is usually the cause (IS portion of ES). This is because an LM caused by the IS portion of a signal appears in both the sum and difference signal spectra at the same frequency. If so, frequency domain processing according to Rule 1 should not be carried out at this frequency range.

For example, in Fig. 18, the sum and difference signal spectra each have two maxima at 2.12 and 2.17 kHz, which rise above the values of the closer environment (approximately 59 dB). It can therefore be assumed that both maxima result from the useful signal. These maxima of the difference signal spectrum may remain unchanged for further signal processing.

In Fig. 20, however, the difference signal spectrum shows several spectral lines with a higher level that are not supported by the spectral lines of the sum signal. It can be deduced, therefore, that the high-level spectral lines of the difference signal originate from an interference signal. A reduction to the level of the sum signal spectrum can be carried out.

Local maxima are identified separately for the sum and difference signal spectra. The median-filtered difference signal spectrum can be seen in Fig. 21(a). A maximum in one of the spectra exists when the spectrum exceeds the median value of the difference signal spectrum for a specified value in decibels.

In Figs. 21(a) and 21(b), LMsum and LMdiff (marker lines near the bottom of the graph) indicate the maxima of the sum and difference signal spectra. If LMsum lies within the bandwidth of LMdiff, a superordinate LM is reported, which causes a level reduction of the difference signal spectrum in the corresponding frequency bandwidth to be blocked. In the case of the example in Fig. 21(b), this is true only for a narrow frequency band at 5.75 kHz. The following can be defined:

- **Rule 3 for signal processing of local level maxima in the frequency domain**.
  If a local or frequency-selective maximum is identified in the difference signal spectrum, Rule 1 for

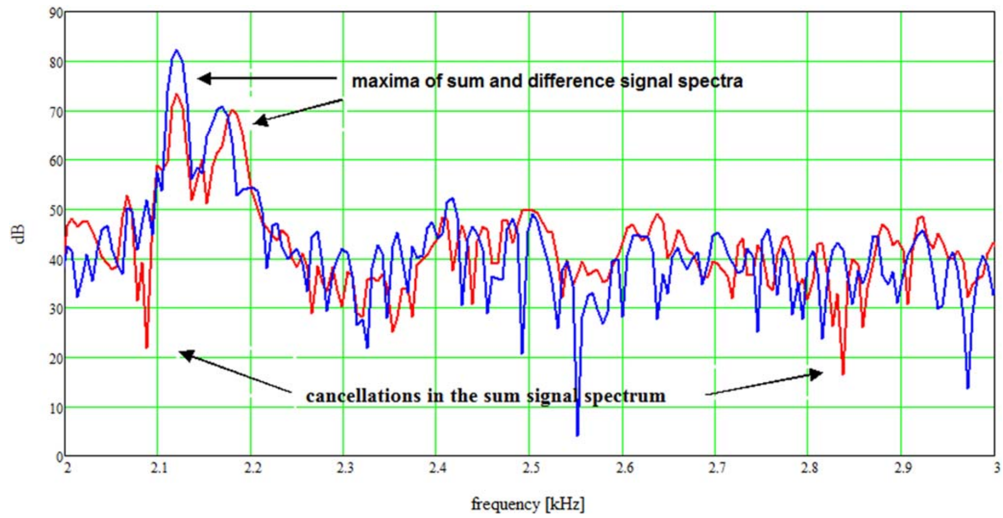

Fig. 18. Difference signal spectrum (blue) and sum signal spectrum (red) in equivalence stereophony (ES) of an undisturbed audio signal [2].

signal processing in the frequency domain is carried out only if the sum signal spectrum does not support this LM within a predefined frequency range by its own LM.

It should be noted that the comb filter effects in the sum and difference signal spectra are more pronounced the lower the IS portion is, especially if the source spectrum is broadband and not a line spectrum. Then minima in the sum signal magnitude spectrum coincide with maxima in the difference signal spectrum and vice versa. Signal processing according to Rules 1–3 reduces interference but may change timbre after dematrixing. This is the subject of further investigation.

2.4 Signal Classification and Temporal Processing

Audio signals can have a transient or stationary character. Transient signals are characterized by an increase in power within the shortest periods of time, often associated with preceding signal pauses or silent passages. Stationary signals have a more continuous power time line.

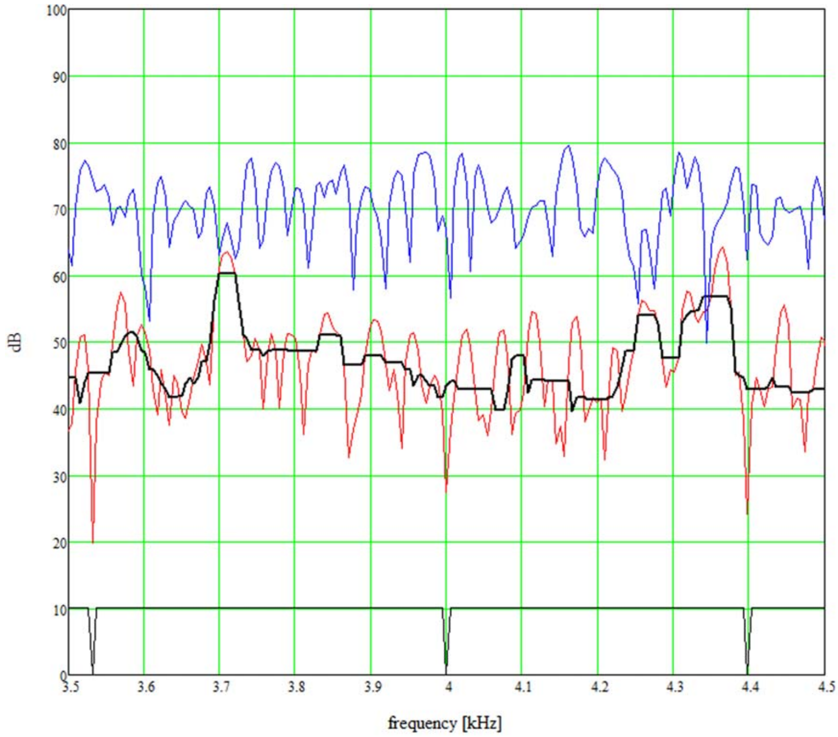

Fig. 19. Disturbed difference signal spectrum (blue), sum signal spectrum (red), median value of the sum signal spectrum (bold black), and marker of cancellations (CANs) in the sum signal spectrum (black).

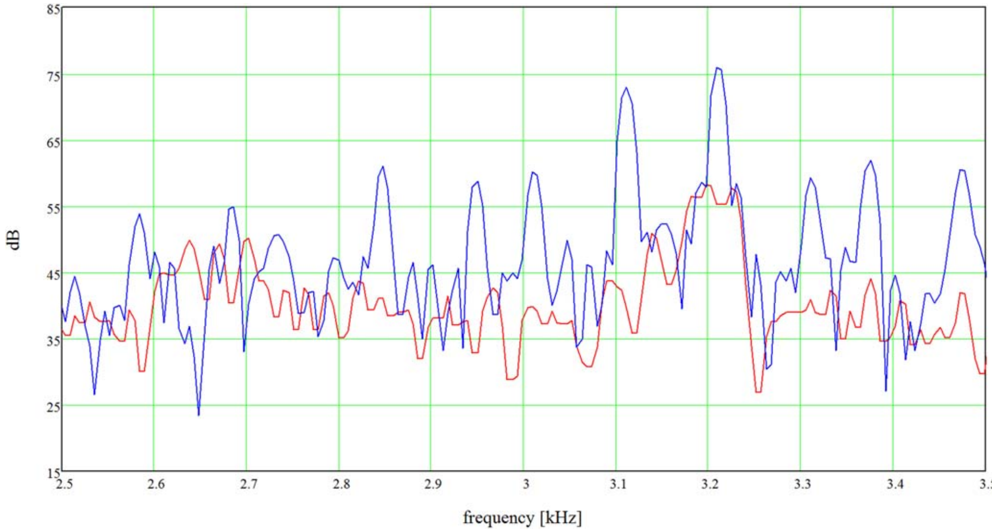

Fig. 20. Sum signal spectrum (red) and difference signal spectrum with frequency-selective interference (blue) [2].

Interfering signals can most effectively be reduced in the frequency domain if the magnitude spectra of the useful signal and that of the interfering signal differ significantly.

Unfortunately this does not apply to transient useful signals (such as guiro and castanets) because they have an almost white spectrum, and there are few differences to the noise as interfering signal. A reduction in selective frequencies can hardly take place. The residual noise is therefore high in such cases. Further disturbances are added: The processing of noise in the frequency domain and transformation back into the time domain causes an alias spread out over the block [19]. It is mostly hidden for stationary useful signals.

As long as the transient useful signal is present, residual noise is masked simultaneously. Residual noise that occurs after a transient can be better masked because the natural transient signals settle slower, and hearing has temporal post-masking. The masking of residual noise, which precedes a transient, is lower [18]. In signal pauses before a transient noise, a so-called pre-echo can be audible [19].

If the useful signal has a transient/impulse character, and noise is superimposed on the difference signal, then the residual noise can be reduced (the pre-echo among others) by an additional signal processing in the time domain (temporal processing). The difference signal is herewith limited to the envelope of the sum signal (clipping). On an example of a transient wave form signal, Figs. 22(a)–22(c) show how additional temporal processing reduces pre-echoes.

If the useful signal is transient and/or short signal pauses are present within the block, then clipping can also help to temporally eliminate or reduce disturbances with a transient character. The limiting to the envelope curve reduces the interfering energy in the case of strong disturbances. In these cases, the magnitude spectrum of the difference signal after clipping lies below that of the untreated magnitude spectrum. The effect of the original disturbance on the magnitude and phase of spectral lines is reduced.

On the other hand, the clipping itself produces noise with a nearly white spectrum, which manifests itself as interferent portions in the magnitude and phase of the difference signal spectrum. As this effect increases, the more signal components are cut off. If it comes to a level increase in spectral lines in this regard, then this can be corrected by the signal processing in the frequency domain. Level reductions cannot be corrected. In the frequency domain, the phase spectrum is further processed unchanged. It is therefore necessary to decide from block to block whether the limiting to the envelope curve is to be applied.

2.4.1 Criterion for the Use of Temporal Processing

If the sum signal (useful signal) is stationary, or it has a temporally continuous signal form, its frequency spectrum usually provides sufficient gaps for a selective interference reduction in the frequency domain of the difference signal. In the case of stationary signals, the temporal processing (limiting to the envelope curve) often worsens the residual noise and thus also the channel separation. Therefore it is better to turn off temporal processing in these cases.

In contrast, if the transients dominate the signal, it is advantageous to additionally use temporal processing. In these instances, the temporal processing reduces the pre-echoes in particular. The following can be defined:

- **Rule 3 for signal processing in the time domain.** Temporal processing (limiting the difference signal to the envelope curve of the sum signal) is turned on, if in the sum signal (useful signal) transients dominate. For identification of transients, the envelope of the sum signal σ is examined. A transient is considered identified if the envelope curve increases by more than x% within a time segment Δt. Determination in percentages allows for level-independent identification.
  Temporal processing takes place before the processing in the frequency domain.

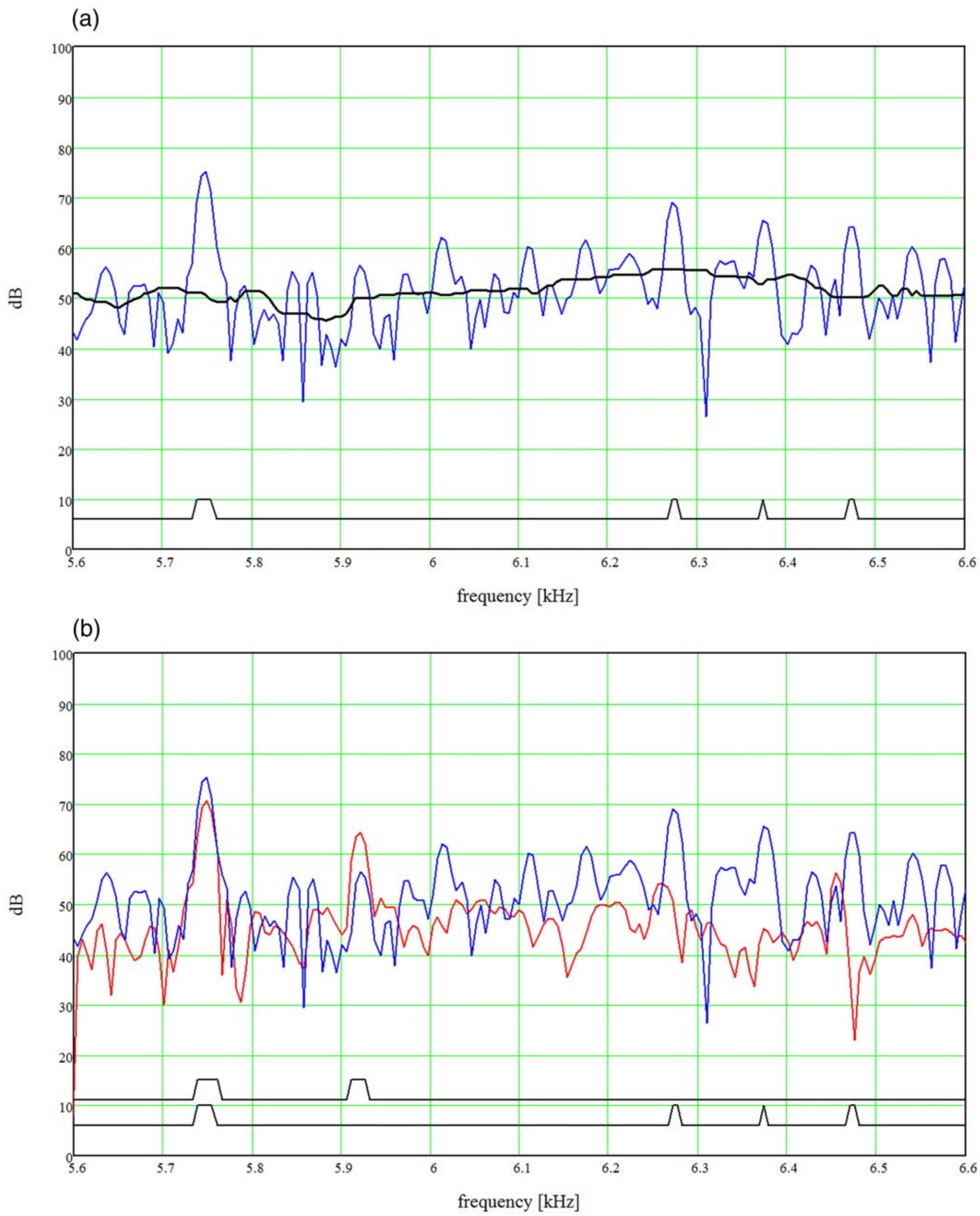

Fig. 21. (a) Identifying local maxima (LM). Difference signal spectrum (blue), median (bold black), and marker line indicating maxima (LM) (black) [2]. (b) Comparing local maxima. Difference signal spectrum (blue), sum signal spectrum (red), and marker lines indicating maxima (LM) in sum and difference signal spectra (black) [2].

The above rules for signal behavior are designed to be followed by all types of sound source signals. Source signals that are not subject to transmission interference are accordingly not audibly changed by the algorithm.

3 SIGNAL PROCESSING OVERVIEW

The signal processing rules are summarized here in the order of their application:

1. In the time domain (TD) IS or ES identified.
2. The envelope of the sum signal is calculated.
3. The envelope is examined for transients (Rule 3 TD).
4. In case of IS and in case of a transient, the difference signal is clipped to the envelope of the sum signal (modified Rule 1 TD and Rule 3 TD).
5. In case of ES and in case of a transient, the difference signal is clipped to the envelope of the sum signal (Rule 2 TD and Rule 3 TD).
6. In ES, CANs and LM are identified in the frequency domain (FD). The difference signal magnitude spectrum is lowered to the median of the sum signal magnitude spectrum (Rule 2 FD) at frequencies with CANs. Further processing (see step 7) of the difference signal magnitude spectrum is blocked in the frequency range of an LM if this LM is supported by an LM of the sum signal magnitude spectrum (Rule 3 FD).

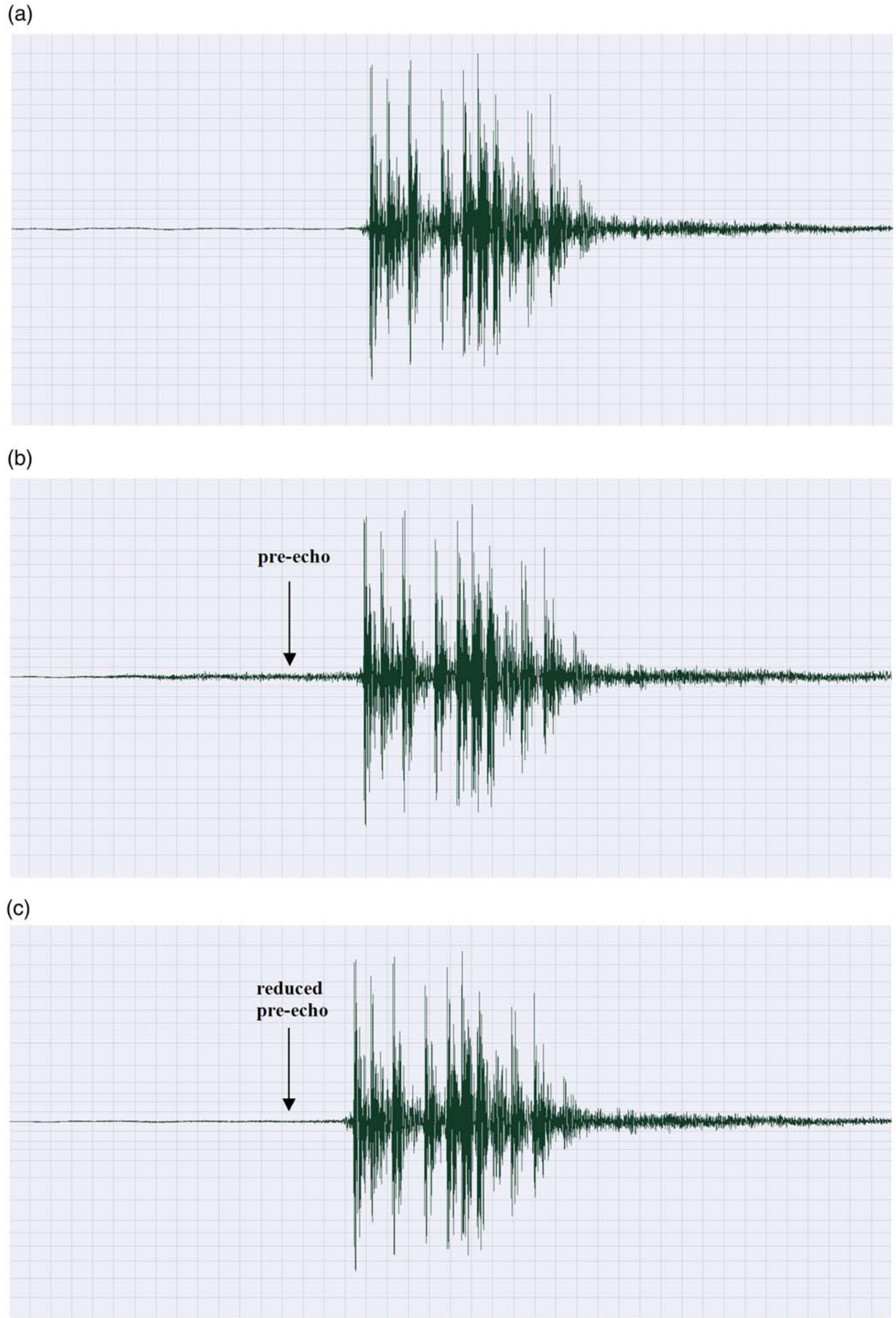

Fig. 22. (a) Undisturbed guiro, without signal processing = original [2]. (b) Noisy guiro, after signal processing in the frequency domain → pre-echo [2]. (c) Noisy guiro, after signal processing in the time and frequency domains → reduction of the pre-echo [2].

7. The difference signal magnitude spectrum is reduced to the sum signal magnitude spectrum (Rule 1 FD), in case of ES at frequencies other than those of CANs and LM.

The signal processing steps for NR are embedded in a weighted-overlap-add structure. The reference implementation (in real time) uses a block length of 4,096 samples with a 50% overlap and systematic latency of 93 ms. A root Hanning weighting function is applied before and after signal processing in the frequency domain. This minimizes spectral leakage and reduces unwanted signal changes at the block boundaries. The weighted-overlap-add structure is signal transparent, i.e., as long as no changes to the signal are made, then the output signal corresponds to the input signal.

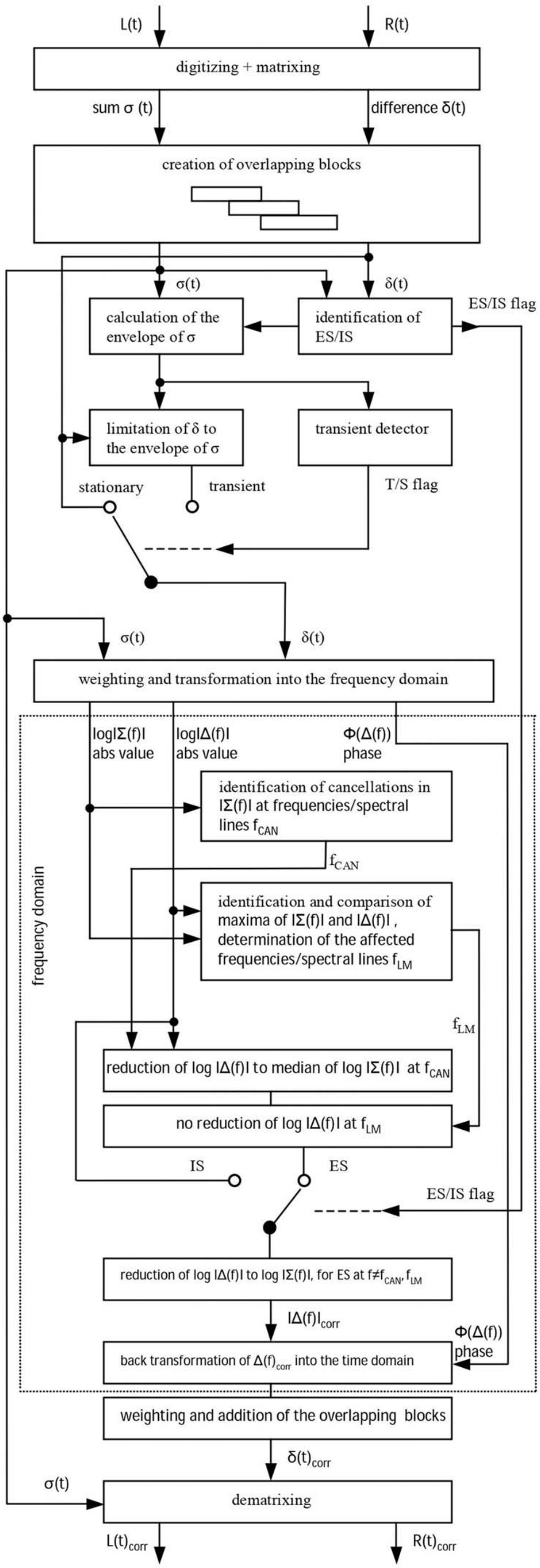

Fig. 23. Block diagram of signal processing [2].

4 NR CAPABILITY OF THE METHOD AND COMPATIBILITY

The proposed method is designed to work above the FM threshold and approximate mono quality regarding the audible stereo audio-to-noise ratio, while preserving the L/R channel separation. NR is up to 20 dB for random FM noise. To fully exploit the benefits of the method, receiver functions such as stereo-blend and High-Cut should be disabled. For a receiver using this technology, the stereo coverage area of a transmitter is adjusted to the mono coverage area.

The method reduces noise and other types of interference that occur in the difference signal and exceed the interference in the sum signal. This includes all interference that occurs on the transmission chain, from matrixing in the stereo encoder to dematrixing in the stereo decoder of the receiver, such as

- Inherent noise of the FM exciter and transmitter;
- Radio transmission interference;
- RF-adjacent channel and co-channel interference;
- Multipath interference;
- Inherent noise in the RF part of the receiver;
- Non-linear distortion products due to the limitation of the intermediate frequency (IF) bandwidth;
- Interference by demodulation of harmonics of the pilot tone signal within the stereo decoder;
- Beat notes from subcarriers (e.g., RDS/RBDS, Subsidiary Communications Authority (SCA);
- Disturbances due to signals of purely digital or hybrid transmission systems, such as an in-band on-channel system (IBOC); and
- Disturbances and crosstalk within hybrid systems, which have an impact on the difference signal of the analog transmission system.

Interference that occurs in the sum channel, i.e., even in the case of pure mono reception, cannot be eliminated by this method.

4.1 Measurement Results

4.1.1 Signal-to-Noise Ratio

Noise measurements were performed without a useful audio signal in three stages. Random noise was recorded from a Hi-Fi receiver connected to an FM transmitter via an attenuator. The receiver has no stereo-blend or High-Cut function. The transmitter was fed with a stereo encoder to produce a pilot tone signal. De-emphasis was 50 μs. The left and right output of the receiver was recorded and then processed. The noise level was related to the level of a sinusoidal signal producing +/–67.5 kHz FM deviation. An SNR value was calculated from this. Table 1 shows the measured stereo SNR before and after processing.

The results show the expected decrease in NR with increasing input SNR. The amount of NR near the FM threshold is close to the theoretical value of the mono-gain of 20.75 dB (50 μs). The mono-gain (see Sec. 1) is the difference between the stereo noise to the mono noise, measured

Table 1. Noise reduction as a function of stereo signal-to-noise ratio (SNR).

Stereo SNR [dB] before processing	Stereo SNR [dB] after processing	Noise reduction [dB]	Remarks
25.8	44.9	19.1	Near FM threshold
35.8	55.1	18.0	
45.8	62.2	16.4	

in decibels. The algorithm approximates the stereo noise to the mono noise.

4.1.2 Stereo Separation

Stereo separation, respectively crosstalk attenuation (CA), was measured at five values of input SNR and with four frequencies. A useful sinusoidal signal with a level corresponding to +/–67.5 kHz FM deviation was added to the left audio channel of the noisy signal. The noisy stereo signal was then processed. CA was measured as the ratio of the useful signal power in the left channel to the power in the right channel in a frequency range of 0–15 kHz. The audio SNR was set by attenuating the RF level. SNR 25.8 dB corresponds approximately to the FM threshold. A further reduction of the RF level would lead to unreliable measured values. The measurements with SNR 15.8 and 5.8 dB were also made at the FM threshold. Instead of lowering the RF level to reduce the SNR, the level of the sinusoidal signal was reduced by 10 and 20 dB in these two cases.

Fig. 24 shows a stereo separation at the FM threshold of about 41 dB for a sinusoidal signal with a +/–67.5 kHz FM deviation (stereo) on only one channel (L or R). CA decreases with a decrease of the desired signal. A sinusoidal signal with a level 20 dB lower (+/–6.75 kHz deviation) has a CA of about 23 dB (see SNR 5.8 dB). This means that at the FM threshold, quieter signals are still reproduced with sufficient channel separation.

At the frequency of the sinusoidal signal, the signal in the other channel has a narrowband noise character (musical noise) and takes on the value of the original noise. Fig. 25(a) shows the spectrum of the left and right stereo channels before NR of a signal with a 25.8 dB stereo SNR. Fig. 25(b) shows the situation after NR.

It should be noted that an extreme case of intensity stereophony is considered here ($R = 0$ or $L = 0$). This case does not occur with ES, because the localization of a sound source by the listener is not achieved by intensity differences alone, but also by time differences between L and R.

4.1.3 Frequency Response

Swept sine measurements confirm the flat frequency response, independent of the SNR.

4.1.4 Nonlinear Distortion

The proposed method is capable of reducing nonlinear distortion products caused by, for example, narrowing the IF bandwidth of the receiver. Since mono demodulation results in lower distortion, the stereo distortion products can be reduced according to the proposed method. Figs. 26(a)–26(c) show the audio spectra of a Hi-Fi receiver after stereo decoding in modes IF wide [Fig. 26(a)], IF narrow [Fig. 26(b)], and IF narrow with processing according to the proposed method [Fig. 26(c)]. The FM transmitter was driven by a stereo encoder with $R = 1$ kHz, resulting in an FM deviation of +/–67.5 kHz. Total FM deviation including pilot tone was +/–75 kHz. Please note that narrowing the IF bandwidth results in higher nonlinear distortion products in the MPX band [see Fig. 26(b) vs. Fig. 26(a)]. Stereo decoding translates these distortion products into the audio frequency band and applies de-emphasis to them.

As Fig. 26(c) shows, signal processing in this example results in a reduction of second-order harmonic distortion of 11.7 dB at 2 kHz, from 1.0% to 0.26%. The third harmonic is reduced by 6 dB at 3 kHz, from 0.14% to 0.07%. The average noise level and nonlinear distortion products in the range above 5 kHz can also be reduced. It it also apparent that the crosstalk of the second harmonic at 2 kHz in the left channel (blue) can be reduced by 31 dB. The stereo separation at the fundamental 1 kHz is not changed by the signal processing.

4.1.5 Compatibility

The presented signal processing method is fully compatible with transmissions according to ITU-R BS.450, in-band on-channel system, RDS/RDBS, single-sideband suppresed-carrier modulation (SSBSC), and Subsidiary Communications Authority (SCA).

5 BEHAVIOR OF THE METHOD IN THE LIMITING CASE

Noticeable changes of the audio signal may occur at very low SNR. Then masking the residual interference in the channels L or R after dematrixing by the useful stereo signal is not effective anymore. Masking takes place in the time, frequency, and spatial domains or any combination thereof. It is dependent on the properties of the stereo signal in these threew domains.

The remaining errors after NR manifest themselves in artifacts such as

(1) Localization blur (human hearing is insensitive),
(2) Reduced stereo separation,
(3) Musical noise, and
(4) Pre and post-echo effects with speech.

The main reason is that interference has changed the phase spectrum of the difference signal at the significant spectral lines.

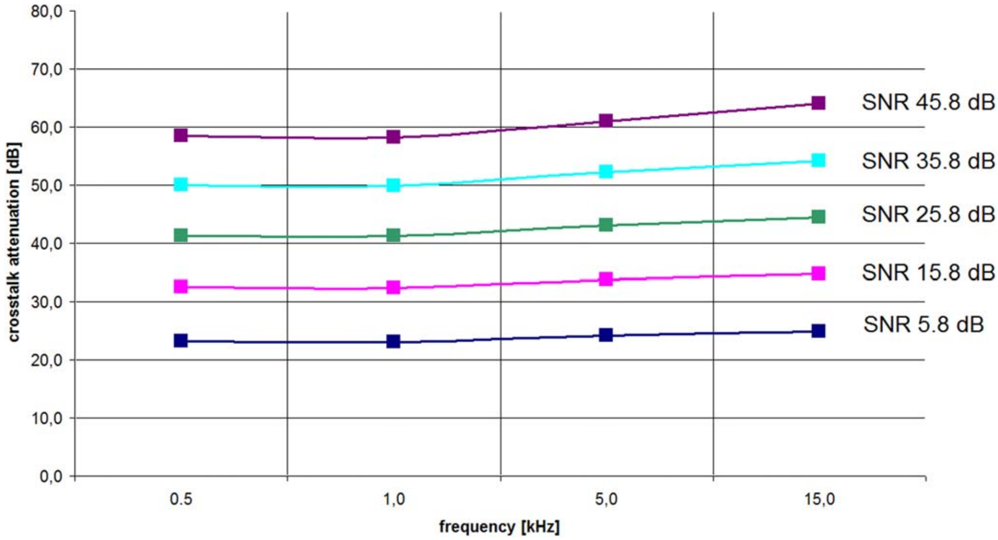

Fig. 24. Crosstalk attenuation as a function of input stereo signal-to-noise ratio (SNR) and frequency.

6 COMPARISON OF THE PROPOSED NR METHOD TO OTHERS IN RECEIVING FM STEREO SIGNALS

The proposed method works on the receiving side. Noise and interference is reduced over a broad range of receiving levels, from the FM threshold to the point where a further increase of receiving level no longer changes the audio SNR. The proposed method relies on the L and R signals of an FM stereo receiver. NR is the result of shaping the audio signal to permissible values according to the established rules of signal behavior. This eliminates the need to estimate the type and magnitude of the noise signal. In the following discussion, the proposed method is compared to five other methods of NR in receiving FM stereo signals.

The FM receiver chip in [8] uses three metrics—a received signal strength indicator, estimated SNR, and estimated multipath interference—combined to gradually control the stereo-blend function with attack and release times for each metric. High-Cut control is employed on audio outputs with degradation of the signal due to low SNR and/or multipath interference. Here the high audio frequencies are attenuated. Both metrics have their own attack and release times. Stereo-blend typically begins at 280 μV. Full mono is typically achieved at 32 μV. Soft mute of the audio signals typically begins at 2 dB SNR (near FM threshold). According to the data sheet, the FM chip is designed for optimum "fidelity," i.e., the stereo fading function can start working early to maintain a high SNR.

The FM receiver chip in [9] uses the field strength input (received signal strength indicator) to control the High-Cut and stereo-blend functions. If excessive noise or multipath interference is detected, the stereo separation is reduced (stereo-blend). High-Cut control is employed on audio outputs with degradation of the signal due to low SNR and/or multipath interference. A multipath condition is detected by evaluation of the field strength signal spectrum in the vicinity of 19 kHz. A noise blanker removes signal spikes, for example, caused by the car ignition, by holding the output signals of the stereo encoder for the blanking time.

[19] describes a system for improving the subjective SNR of a received FM stereo signal without changes of the frequency response. NR is achieved by gradually blending from stereo to mono as a function of program level. Adjusting attack and release times of the program level-dependent blend control and restricting the blend to frequencies above 5 kHz avoid noise pumping and recognizable changes in separation. The author refers to some basic psychoacoustic principles, which the proposed NR is taking advantage of:

- Loud sounds tend to mask sounds of lower level,
- At low sound levels the directional perception is not as acute,
- Low frequency sounds do not mask high frequency sounds very effectively,
- Sounds are localized primarily by dominant spectral components, and
- The brain localizes by the first few milliseconds of the sound.

No gross changes in image localization are claimed. The NR works most impressively with moderately quiet stations. The amount of NR has not been determined.

In [20], an NR algorithm is presented that covers at least four quality degrading effects:

- Increased level of hissing noise,
- Reduced stereo separation to keep the noise low,
- Appearance of noise pulses below the FM threshold, and
- Short but loud bursts of noise with RF signal loss (below FM threshold).

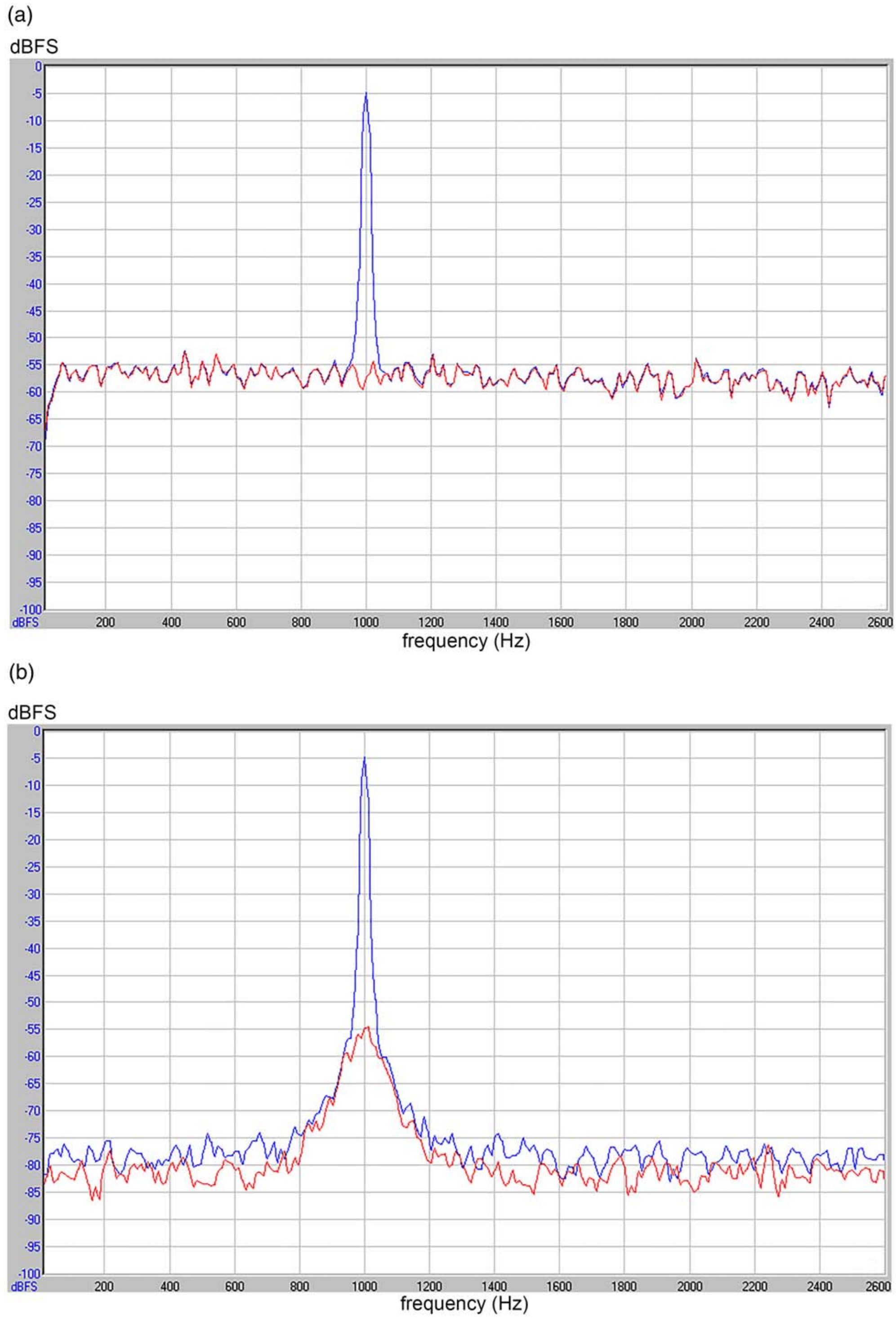

Fig. 25. (a) Signal spectrum (averaged) before noise reduction. (b) Signal spectrum (averaged) after noise reduction with musical noise in the right channel (red) as an artifact of noise reduction. The measured value for the crosstalk attenuation is 41 dB.

For a comparison, only the first two points are discussed, since the proposed method in this paper is designed to work above the FM threshold.

The algorithm reduces noise mono and stereo noise. Since noise measurements in the sum and difference signals are not reliable, the algorithm relies on special circuitry in the IF and MPX part of the receiver used for noise estimation. NR is based on taking a noise sample in the quadrature demodulated (L–R) signal or in a narrow frequency band around the pilot frequency. The noise floors in the difference and sum signals are then extrapolated from the noise samples. The actual NR is done in overlapping blocks of 256 audio samples. A Fast Fourier Transform (FFT) calculates the spectra of the sum and difference signals. The local SNR at a frequency bin is calculated by dividing the bin signal power by the extrapolated noise power. The signal

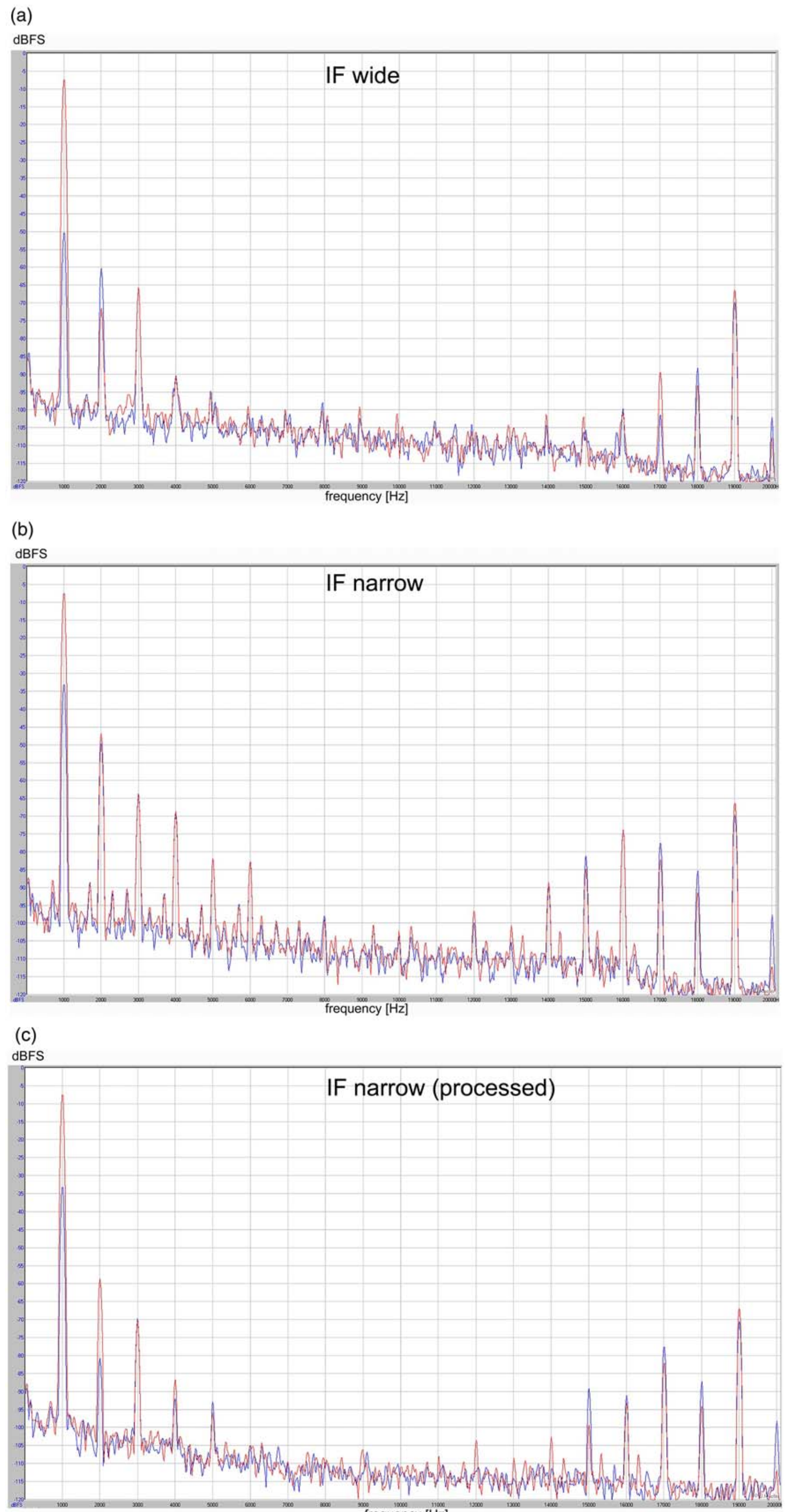

Fig. 26. (a) intermediate frequency (IF) wide. (b) IF narrow. (c) IF narrow (processed).

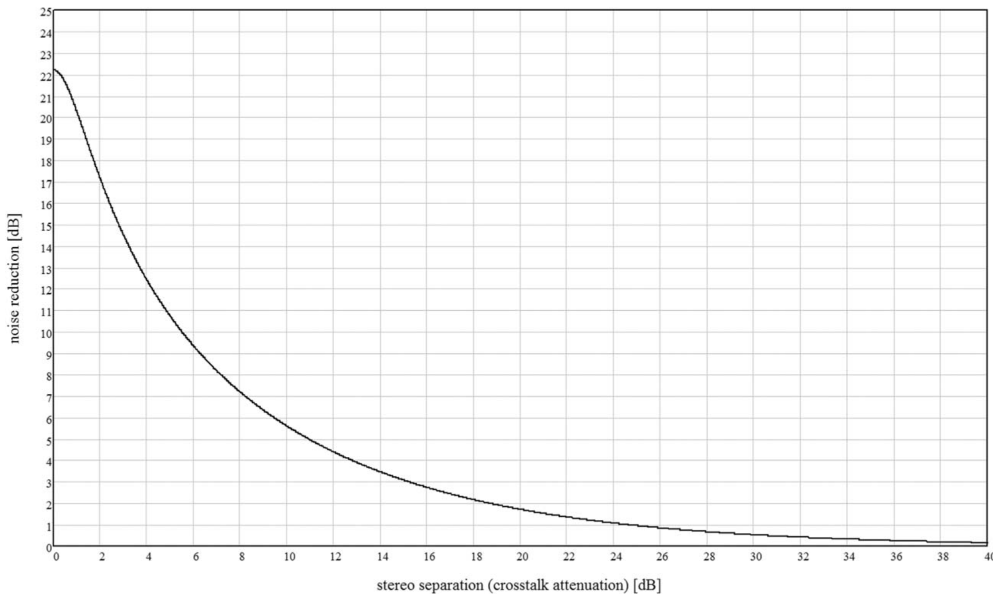

Fig. 27. Noise reduction at the FM threshold as a function of stereo separation with 75-μs de-emphasis for a frequency flat stereo-blending.

power at a frequency bin is then reduced as a function of the calculated SNR at this bin. A low SNR leads to a greater reduction in bin level. This results in some distortion of the desired signal depending on the instantaneous power at that bin.

Since the extrapolation of a narrow band noise sample to the entire audio frequency band is tailored to a certain noise power density function (presumably white RF noise), noise signals with line spectra cannot be meaningfully extrapolated. Even when the noise sample is taken from the quadrature signal over a range of 23–53 kHz, the line spectrum does not match the in-phase noise spectrum. This indicates that the algorithm, at the very least, distorts the useful signal spectrum in this case.

Furthermore weighted NR [20] does not take into account the stereo base limit, as is the case with the method proposed here. Therefore level reduction of the difference signal at certain frequencies, depending on the calculated noise, may exceed the value of the sum signal and degrade channel separation. Stereo separation is about 15 dB worse compared with the method presented in this paper. The method according to [20] achieves a 60-dB SNR over the entire RF signal level range, presumably by supporting NR in the mono signal. This gives the method a 15-dB (5 dB) advantage at an RF signal level of –100 dBm (–110 dBm) over the proposed method in this paper.

In [21], a method for improved pilot tone recovery and stereo decoding is presented. A Costas loop, which evaluates the DSBSC-modulated difference signal, is used to compensate for phase differences between the recovered pilot tone and DSBSC-modulated difference signal. Such phase errors limit the amount of stereo separation. They may result from an uneven phase response in the receiver MPX band or allowable phase difference in the stereo encoder at the transmitter site. Stereo separation can be held high even in poor reception conditions. According to the patent description, the method achieves a stereo separation of about 45 dB at a CNR of 12 dB (RF bandwidth 200 kHz). At this point, the stereo SNR is given by about 27 dB. The patent also covers a stereo-blend circuit. It can be used to improve SNR at the expense of stereo separation. The method [21] works in the MPX band after the FM demodulator. The achieved values of the channel separation must be seen in connection with the stereo SNR. The method does not allow to keep both values high. In order to improve the stereo SNR, the channel separation must also be reduced here via the stereo-blend function.

The following will discuss and compare the methods in their application above the FM threshold, since this is the working range of the proposed method. Therefore strategies to improve mono reception (such as a noise blanker) are not covered by the comparison. The types of reducible noise will also be examined.

1. [8], [9], and [21] use frequency flat stereo-blending to reduce noise at the expense of stereo channel separation. [19] restricts stereo-blend and NR to above 5 kHz
2. [8] and [9] apply stereo-blend and High-Cut to reduce multipath interference. It is up to the device manufacturer to define start and end points for stereo-blend in the programming of the chip.
3. NR in [8], [9], and [20] is based on noise estimation. While random noise can be estimated in [8] and [9] by use of the RF signal strength, [20] uses noise samples in the Q-channel or around the pilot tone frequency.
4. [19] reduces any kind of noise above 5 kHz. NR varies with program level. Since most of the noise of the difference signal is below 5 kHz under weak reception conditions, the NR is only moderate.
5. [8], [9], and [19] apply attack and release times to the blend function to reduce noise pumping and achieve a gradual transition.

Table 2. Maximum achievable noise reduction (at the FM threshold) as a function of stereo separation and stereo base width for a frequency flat stereo-blending.

Stereo separation (CA)	Stereo base width	Noise reduction (NR)
9.5 dB	75%	6 dB
6.7 dB	50%	8.5 dB
4 dB	25%	12 dB

6. [21] improves channel separation, even in the case of misaligned stereo separation at the transmitter site, at the expense of audio SNR. Stereo-blend is used to improve SNR. Stereo separation is claimed to be 45 dB with an SNR of 27 dB at the FM threshold.
7. [20] reduces noise in the sum and difference signals at each frequency bin after FFT, depending on an estimated SNR at that bin and based on an assumed RF noise spectrum shape (e.g., white noise). The single channel noise suppression algorithm used leads to a high NR at the expense of a distortion of the useful signal spectrum. Lowering the levels of the sum and difference signal frequency bins (depending on the respective SNR), relying solely on the estimated SNR, leads to a stereo separation that is worse than the proposed method. Furthermore interference with a line spectrum cannot be identified as such or be reduced while maintaining stereo separation. Multipath distortion is not fully captured by the NR algorithm because the SNR estimation is not designed for it.

The stereo-blend function will now be examined. Fig. 27 and Table 2 show the dependence of NR and CA for a frequency flat stereo-blending, which applies to all methods except [19] and [20]. The corresponding calculation can be found in APPENDIX A. For a given CA, NR of [19] is lower than shown in both Fig. 27 and Table 2 because it takes place only above 5 kHz.

Stereo separation can be related to stereo base width according to Fig. 1 in [10]. Set stereo separation (CA) to be equal to interchannel level difference to get the phantom source shift = % stereo base width.

From Fig. 27 and Table 2, it can be seen that for a noticeable NR, the stereo base width has to be significantly narrowed. The proposed method, on the other hand, achieves a stereo separation of >40 dB (100% stereo base width) even with an NR of 19 dB.

The method in [20] uses a single channel noise suppression algorithm in the sum and difference signals. It leads to a high NR for random noise at the expense of a distortion of the useful signal. For comparison, Table 3 shows the SNR improvement and stereo separation of [20] vs. the proposed method. Additionally, stereo separation is given for typical legacy devices according to [20].

Table 3. signal-to-noise ratio (SNR) and stereo separation or crosstalk attenuation (CA).

Input SNR	RF signal level [20]	Output SNR (proposed method)	Output SNR [20]	CA (proposed method)	CA [20]	CA (legacy FM devices)
25.8 dB	–110 dBm	44.9 dB	60 dB	>40 dB	26 dB	0 dB
35.8 dB	–100 dBm	55.1 dB	60 dB	>50 dB	35 dB	1 dB
45.8 dB	–90 dBm	62.2 dB	60 dB	>58 dB	40 dB	5 dB

Table 4. Comparison of the noise-reduction methods.

	[8]	[9]	[19]	[20]	[21]	Proposed method
NR by reducing the stereo base (stereo-blend)	√	√	√	-	√	-
Single channel noise suppression in the sum and difference signal	-	-	-	√	-	-
NR by reducing excessive stereo base width	-	-	-	-	-	√
NR, frequency-selective	-	-	√	√	-	√
NR, independent on program level	√	√	-	-	√	√
NR, independent on estimated SNR based on external criteria	-	-	√	-	√	√
NR with flat frequency response	-	-	√	√	√	√
Reduction of noise with line spectrum without reducing the stereo base	-	-	-	-	-	√
Reduction of non-linear distortion products without reducing the stereo base	-	-	-	-	-	√
Reduction of multipath distortion without reducing the stereo base	-	-	-	-	-	√

While in [20] the emphasis is on SNR at the expense of signal distortion and stereo separation, the proposed method achieves high stereo separation down to the FM threshold without signal distortion but with an SNR lower than that in [20]. Solely the proposed method with its CA of >40 dB assures sufficient stereo separation with quiet passages even at weak reception conditions.

With respect to interference types other than random noise, and in contrast to the proposed method, none of the above methods is able to reduce multipath distortion and nonlinear distortion products while maintaining stereo separation. Additionally, interference with a line spectrum (e.g., machine noise) cannot be reduced without restricting the stereo base.

The proposed method compares the sum signal spectrum with the difference signal spectrum at each spectral line (each bin after FFT). Lowering the level of bins while respecting the stereo base does not distort the useful signal. According to the derived rules, any kind of interference can be reduced to a quality comparable to mono while respecting the stereo separation. This includes, but is not limited to, impulse noise, non-linear distortion products in the difference signal that exceed the values of the sum signal, and interference with a line spectrum, such as machine noise. NR is not dependent on program level or an estimated SNR (based on external criteria).

Unique features of the proposed method include

- L/R input that allows processing after FM reception or after recording,
- Reduction of multipath interference without restricting the stereo base,
- Reduction of machine noise with line spectrum without restricting the stereo base, and
- Reduction of non-linear distortion products exceeding the values of the sum signal.

Audio examples processed according to the proposed method can be heard and downloaded at www.fm-stereo.com.

7 CONCLUSION

An NR method is proposed that reduces noise and interference of a received FM stereo signal to a quality comparable to mono while preserving the stereo separation and frequency response. The method works over the whole range of reception levels down to the FM threshold. Since only L and R are used as input signals, the method can also be used after a recording. The method is applicable worldwide with FM stereo transmissions according to the pilot tone system described in ITU-R BS.450. For a receiver using this method, the stereo coverage area of a transmitter is adjusted to the mono coverage area.

The signal processing rules are directly related to rules for the behavior of undisturbed signals, derived by considering the matrixed signals of common recording techniques. Signal processing addresses the difference signal (L–R) and takes place in the time and frequency domains. According to the signal processing rules in the frequency domain, the magnitude spectrum of the difference signal is reduced to the value of the sum signal, thereby eliminating the excessive width of the stereo base caused by noise. This represents the main part of the NR. The associated phase spectrum is left unchanged, as is the sum signal. Processing results in a noise-reduced stereo signal with maintained stereo separation and frequency response.

The rules for signal behavior are designed to be followed by all types of sound source signals. Source signals that are not subject to transmission interference accordingly are not audibly changed by the algorithm.

8 REFERENCES

[1] ITU-R, “Transmission Standards for FM Sound Broadcasting at VHF,” *Recommendation ITU-R BS.450-4* (2019 Oct.). https://www.itu.int/dms_pubrec/itu-r/rec/bs/R-REC-BS.450-4-201910-I!!PDF-E.pdf.

[2] H.-P. Hahn, “Method for Processing an FM Stereo Signal,” US Patent 10,003,422 (2018 Jun.).

[3] S. Haykin, *Communication Systems* (Wiley, New York, NY, 1994), 3rd ed.

[4] J. Dunlop and D. G. Smith, *Telecommunications Engineering* (CRC Press, Boca Raton, FL, 1994), 3rd ed.

[5] W. D. Gregg, *Analog & Digital Communications: Concepts, Systems, Applications and Services in Electrical Dissemination of Aural, Visual and Data Information* (Wiley, New York, NY, 1977).

[6] R. Shatara, “Device and Method for FM Demodulation With Threshold Extension,” US Patent 9,455,853 (2016 Sep.).

[7] McIntosh Laboratory, Inc., “MR78 AM/FM Tuner Owner’s Manual,” https://www.mcintoshlabs.com/products/tuners/MR87 (2009).

[8] Silicon Laboratories, “Si4704/05-C40: Broadcast FM Radio Receiver for Consumer,” https://pdf1.alldatasheet.com/datasheet-pdf/view/538052/SILABS/SI4704.html (2009).

[9] STMicroelectronics, “TDA7540B: AM/FM Car Radio Tuner IC With Stereo Decoder and Intelligent Selectivity System (ISS),” https://pdf1.alldatasheet.com/datasheet-pdf/view/154311/STMICROELECTRONICS/TDA7540B.html (2006 May).

[10] H. Wittek and G. Theile, “The Recording Angle - Based on Localisation Curves,” presented at the *112th Convention of the Audio Engineering Society* (2002 Apr.), paper 5568.

[11] E. Sengspiel, “Lokalisationsschärfe der Hörereignisse auf der Lautsprecherbasis bei verschiedenen Stereophonie-Aufnahmeverfahren,” http://www.sengpielaudio.com/Lokalisationsschaerfe.pdf (1994 Jul.).

[12] DPA Microphones, “Stereo Recording Techniques & Setup,” https://www.dpamicrophones.com/mic-university/stereo-recording-techniques-and-setups (2019).

[13] J. C. Makous and J. C. Middlebrooks, “Two-Dimensional Sound Localization by Human Listeners,” *J.*

Acoust. Soc. Am., vol. 87 no. 5, pp. 2188–2200 (1990 May). https://doi.org/10.1121/1.399186.

[14] M. Brückner, *Äquivalenzstereomikrofonie - Vergleiche und neue Methoden, Diplomarbeit*, Diploma Thesis, SAE Institute, Frankfurt am Main, Germany (2001 Aug.). http://www.marcousticaudio.de/Stuff/Diplomarbeit.pdf.

[15] R. Streicher, "The Decca Tree—It's Not Just for Stereo Any More," http://www.123seminarsonly.com/Seminar-Reports/2016-01/92463083-Surround-Sound-Decca-Tree-urtext.pdf (2003 Feb.)

[16] F. A. Everest, *The Master Handbook of Acoustics* (McGraw-Hill, New York, NY, 2001), 4th ed.

[17] S. P. Lipshitz, "Stereo Microphone Techniques: Are the Purists Wrong?" *J. Audio Eng. Soc.*, vol. 34, no. 9, pp. 716–744 (1986 Sep.).

[18] A. Spanias, T. Painter, and V. Atti, *Audio Signal Processing and Coding* (Wiley, Hoboken, NJ, 2007).

[19] R. C. Cabot, "A Dynamic Noise Reducer for Sum-Difference Multiplex Systems," *J. Audio Eng. Soc.*, vol. 25, no. 3, pp. 95–98 (1977 Mar.).

[20] J.-H. Chen, T. Baker, E. McCarthy, and J. Thyssen, "System Architectures and Digital Signal Processing Algorithms for Enhancing the Output of Stereo FM Broadcast Receivers," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 226–230 (South Brisbane, Australia) (2015 Apr.). https://doi.org/10.1109/ICASSP.2015.7177965.

[21] G. Lerner, N. Tal, and O. Friedman, "FM Stereo Decoder Incorporating Costas Loop Pilot to Stereo Component Phase Correction," US Patent 7,787,630 (2010 Aug.).

A.1

CALCULATION OF THE MONO AND STEREO NOISE POWER SPECTRAL DENSITIES AND THE MONO-GAIN

The MPX signal (see Fig. 2) has a noise power spectral density proportional to f^2. In the following, the noise power densities are given for the three frequency bands: the baseband (sum signal), lower sideband of the modulated difference signal, and upper sideband. De-emphasis is applied in each case in the baseband (denominator) with a time constant τ of 75 or 50 μs. A factor k is used for scaling purposes as in Fig. 4.

$NPSD_{mono}(f)$ is the noise power density of the baseband (0–15 kHz) when the receiver (without inherent noise after the stereo decoder) is in mono mode [W/Hz].

$$NPSD_{mono}(f) = \frac{kf^2}{1+(2\pi f\tau)^2}. \tag{23}$$

After demodulation by the stereo decoder, the sidebands appear in the baseband.

The lower sideband is mirrored against 38 kHz and shifted by 38 kHz to the baseband.

$$NPSD_{lsb}(f) = k\frac{(38000-f)^2}{1+(2\pi f\tau)^2}. \tag{24}$$

The upper sideband is shifted to the baseband.

$$NPSD_{usb}(f) = k\frac{(f+38000)^2}{1+(2\pi f\tau)^2}. \tag{25}$$

$NSDP_{stereo}(f)$ is the noise power density when the receiver (without inherent noise after the stereodecoder) is in stereo mode [W/Hz]. Because of the dematrixing rule, the baseband noise adds to the noise of the lower and upper sidebands.

$$NPSD_{stereo}(f) = NPSD_{mono}(f) + NPSD_{lsb}(f) + NPSD_{usb}(f), \tag{26}$$

$$NPSD_{stereo}(f) = k\frac{f^2+(38000-f)^2+(f+38000)^2}{1+(2\pi f\tau)^2}, \tag{27}$$

$$N_{mono} = \int_0^{15000} \frac{kf^2}{1+(2\pi f\tau)^2}df, \tag{28}$$

$$N_{stereo} = \int_0^{15000} k\frac{f^2+(38000-f)^2+(f+38000)^2}{1+(2\pi f\tau)^2}df, \tag{29}$$

$NPSD_{mono}(f)$ and $NPSD_{stereo}(f)$ are shown in the logarithmic form in Fig. 4. $NPSD_{lsb}(f)$ is the noise power spectral density of the lower sideband (23–38 kHz) [W/Hz]. $NPSD_{usb}(f)$ is the noise power spectral density of the upper sideband (38–53 kHz) [W/Hz]. N_{mono} is the noise power of the baseband (0–15 kHz) [W]. N_{stereo} is the total noise power of the baseband, lower sideband (23–38 kHz), and upper sideband (38–53 kHz) [W].

The frequency-dependent mono-gain is defined as

$$monogain(f) = \frac{NPSD_{stereo}(f)}{NPSD_{mono}(f)}. \tag{30}$$

It is shown in its logarithmic form in Fig. 5.

The mono-gain [dB] is the difference in noise power when the receiver is switched from stereo to mono.

$$monogain = 10\log\left(\frac{N_{stereo}}{N_{mono}}\right). \tag{31}$$

For a time constant τ of 75 μs (50 μs), the mono-gain is 22.2 dB (20.75 dB).

CALCULATION OF THE INTERDEPENDENCE BETWEEN NR, CA/STEREO SEPARATION, AND STEREO BASE WIDTH

Random RF noise, which translates into a frequency proportional noise voltage after FM demodulation in a receiver, is considered. After stereo decoding, the stereo noise (in channel L or R) is above the mono noise. The difference in decibels is called the frequency-independent mono-gain. It varies with RF signal level and is highest at the FM threshold, with a value of 22.2 dB (with 75 μs de-emphasis).

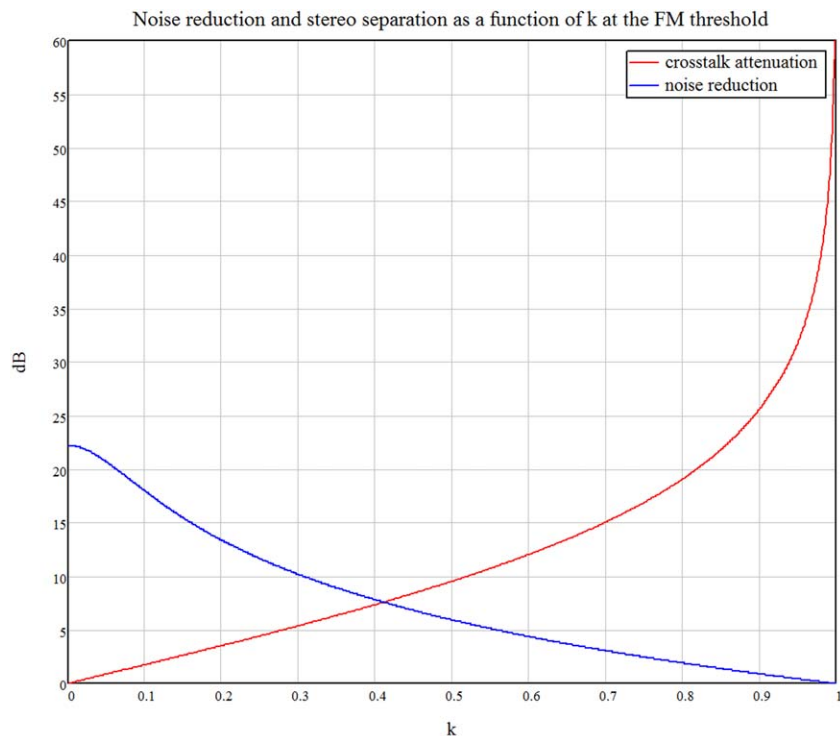

Fig. 28. Crosstalk attenuation (CA) and noise reduction (NR) as a function of the blend factor k at the FM threshold with a mono-gain of 22.2 dB.

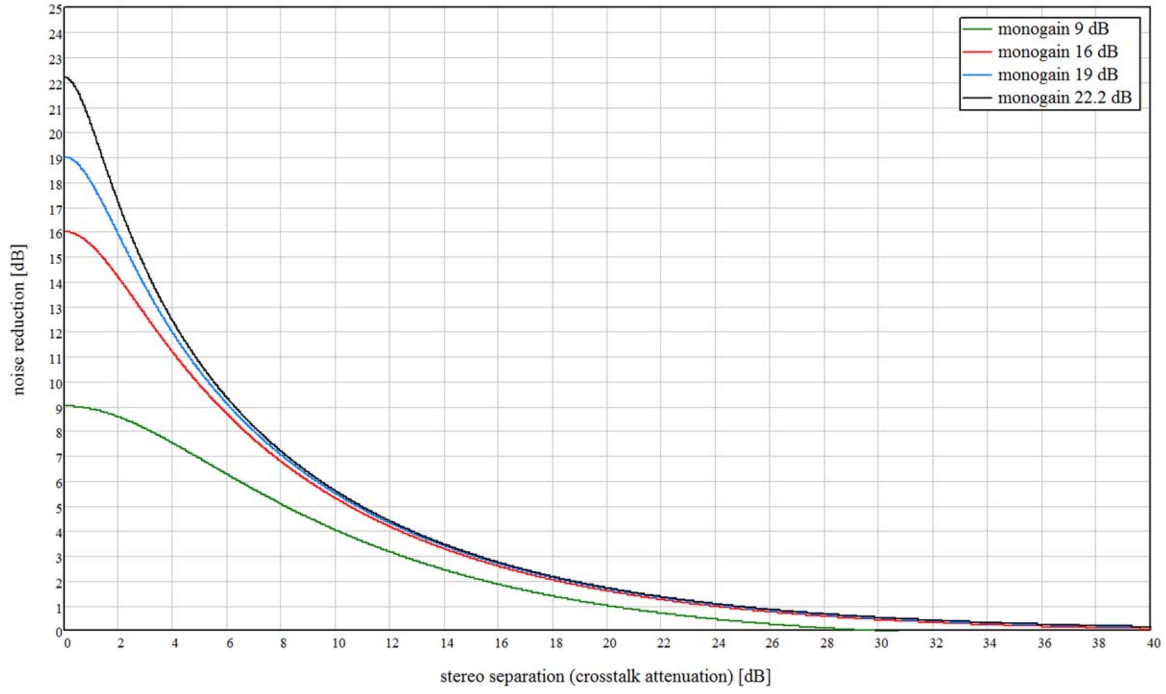

Fig. 29. Noise reduction (NR) as a function of crosstalk attenuation (CA) for different values of the mono-gain.

Table 5. Relationship between stereo separation, stereo base width and NR at the FM threshold.

Stereo separation	Stereo base width	Noise reduction
9.5 dB	75%	6 dB
6.7 dB	50%	8.5 dB
4 dB	25%	12 dB
0 dB	0%	22.2 dB

The frequency-independent mono-gain represents the NR obtained when stereo is switched to mono at the expense of a total loss of stereo separation.

The noise in L and R can be reduced by any value from 0 dB to the mono-gain by reducing the level of the difference signal at the expense of stereo separation.

The dependence of the NR on the stereo separation or CA is formulated. Regarding CA, the receiver is assumed to be ideal (CA = infinite). For the sake of simplicity, white noise will further be assumed in the sum and difference signals after de-emphasis, with a noise power in the sum signal = NP_{sum} and noise power in the difference signal = NP_{diff}.

To be able to evaluate CA or stereo separation, the noise-free source signals L = signal and R = 0 were chosen.

After matrixing in the stereo encoder and adding random noise, the stereo decoder processes the signals:

$$sum = \frac{L+R}{2} = \frac{signal}{2} + \sqrt{NP_{sum}}, \tag{32}$$

$$diff = \frac{L-R}{2} = \frac{signal}{2} + \sqrt{NP_{diff}}. \tag{33}$$

After dematrixing:

$$L = sum + diff = signal + \sqrt{NP_{sum}} + \sqrt{NP_{diff}}, \tag{34}$$

$$R = sum - diff = \sqrt{NP_{sum}} - \sqrt{NP_{diff}}. \tag{35}$$

Introducing frequency flat stereo-blending with the blend factor k ranging from 0 (mono) to 1 (full stereo):

$$sum = \frac{L+R}{2} = \frac{signal}{2} + \sqrt{NP_{sum}}, \tag{36}$$

$$diff = \frac{L-R}{2} = k\frac{signal}{2} + k\sqrt{NP_{diff}}. \tag{37}$$

After dematrixing:

$$L = sum + diff = \frac{signal}{2} + \sqrt{NP_{sum}} + k\frac{signal}{2} + k\sqrt{NP_{diff}}, \tag{38}$$

$$R = sum - diff = \frac{signal}{2} + \sqrt{NP_{sum}} - k\frac{signal}{2} - k\sqrt{NP_{diff}}, \tag{39}$$

$$L = \frac{1+k}{2}signal + \sqrt{NP_{sum}} + k\sqrt{NP_{diff}}, \tag{40}$$

$$R = \frac{1-k}{2}signal + \sqrt{NP_{sum}} - k\sqrt{NP_{diff}}, \tag{41}$$

$$L = \frac{1+k}{2}signal + \sqrt{NP_{sum}} + \sqrt{k^2 NP_{diff}}, \tag{42}$$

$$R = \frac{1-k}{2}signal + \sqrt{NP_{sum}} - \sqrt{k^2 NP_{diff}}. \tag{43}$$

For values of the blend factor k between 0 and 1 and for values of the mono-gain between 0 dB and 22.2 dB, the following equations apply:

CA

$$CA = \frac{L}{R}, \tag{44}$$

$$CA(k) = \frac{1+k}{1-k}, \tag{45}$$

$$ca(k) = 20\log\{CA(k)\} \text{ (crosstalk attenuation in dB)}. \tag{46}$$

NR

The noise of the sum and difference signals is uncorrelated. The noise powers add up after dematrixing and de-emphasis:

$$NP_{total} = NP_{sum} + k^2 NP_{diff}. \tag{47}$$

With k = 0, the receiver was set to mono. NR is then equal to the mono-gain (e.g., 22.2 dB at the FM threshold). With k = 1, there is full stereo with no NR.

$$NR = monogain - 10\log\left(\frac{NP_{total}}{NP_{diff}}\right) \text{ [dB]}, \tag{48}$$

$$NR(k) = monogain - 10\log\left(1 + nk^2\right), \tag{49}$$

with the frequency-independent mono-gain:

$$monogain = N_{stereo} - N_{mono} \text{ [dB]}, \tag{50}$$

$$n = 10^{\frac{monogain}{10}} \text{ (in linear measure)}. \tag{51}$$

In the example of Fig. 7 (quieting curves of an FM receiver), it is derived:

- Mono-gain = 9 dB and n = 8 at100 μV;
- Mono-gain = 16 dB and n = 40 at 30 μV;
- Mono-gain = 19 dB and n = 79 at 10 μV; and
- Mono-gain = 22.2 dB and n = 166 at 1 μV (corresponding to the FM threshold).

From Eqs. (1) and (2), it is derived:

$$k(ca) = \left(\frac{10^{\frac{ca}{20}} - 1}{10^{\frac{ca}{20}} + 1}\right), \tag{52}$$

Inserted into Eq. (11), this gives:

$$NR(ca) = monogain, \\ -10\log\left[1 + 10^{\frac{monogain}{10}}\left(\frac{10^{\frac{ca}{20}} - 1}{10^{\frac{ca}{20}} + 1}\right)\right]. \tag{53}$$

Stereo separation can be related to stereo base width according to Fig. 1 in [12]. Set stereo separation (CA) to be equal to interchannel level difference to get the phantom source shift = % stereo base width. Table 5 shows the dependencies for an FM reception at the FM threshold.

THE AUTHOR

Hans-Peter Hahn

Hans-Peter Hahn earned his Dipl. Ing. Degree from the Technische Hochschule Darmstadt (now the Technical University of Darmstadt) in electrical and communications engineering. During his more than 30 years of employment at the broadcaster Hessischer Rundfunk in Frankfurt, he worked in the fields of communications and transmission theory and technology: TV, radio, satellite ground stations, data transmission, micro wave links, forward error correction, measurement technology, acoustics, and analog and digital signal processing and transmission. As the sole proprietor and owner of an engineering business, he develops customized solutions in the field of signal processing.