**S**pecial Topic

**Noise Feedback Coding Revisited: Refurbished Legacy Codecs and New Coding Models**
Stéphane Ragot, Balázs Kövesi, and Alain Le Guyader

# Noise Feedback Coding Revisited: Refurbished Legacy Codecs and New Coding Models

*Stéphane Ragot, Balázs Kövesi, and Alain Le Guyader*
(audiOvisual and sPEech foR quAlity (OPERA) Lab., Orange Labs, Lannion, France)

**Abstract**

Noise feedback coding (NFC) has attracted renewed interest with the recent standardization of backward−compatible enhancements for ITU−T G.711 and G.722. It has also been revisited with the emergence of proprietary speech codecs, such as BV16, BV32, and SILK, that have structures different from CELP coding. In this article, we review NFC and describe a novel coding technique that optimally shapes coding noise in embedded pulse−code modulation (PCM) and embedded adaptive differential PCM (ADPCM). We describe how this new technique was incorporated into the recent ITU−T G.711.1, G.711 App. III, and G.722 Annex B (G.722B) speech−coding standards.

**Keywords**

speech coding; noise shaping; noise feedback coding; G.711; G.722

## 1 Introduction

Noise shaping is a key technique for improving sound quality in telecommunications and multimedia. Analog noise reduction systems relied on companding/expanding in frequency bands to reduce recording noise from magnetic tapes [1]. Pulse−code modulation (PCM) in ITU−T G.711 uses companding/expanding on the time domain magnitude to achieve a near−constant signal−to−noise ratio (SNR) over a large range of input levels [2]. Even if PCM does not involve any noise shaping in the frequency domain, it exploits the human perception of an audio signal on a log scale.

In this paper, we describe noise feedback coding (NFC), a well−known noise shaping technique [3]−[5], [6], [7]. The goal of NFC is to modify the input signal of scalar quantization by feeding back the filtered coding noise. NFC was introduced to shape the spectrum of the quantization noise in sample−based waveform coders such as PCM coders. A similar principle is used in sigma−delta analog−to−digital (A/D) converters, except that the filter is in the feedforward path [8] instead of the feedback path. NFC has attracted renewed interest with the recent standardization of backward−compatible enhancements to ITU−T G.711 and G.722 [9], [10]. It has also been revisited with the emergence
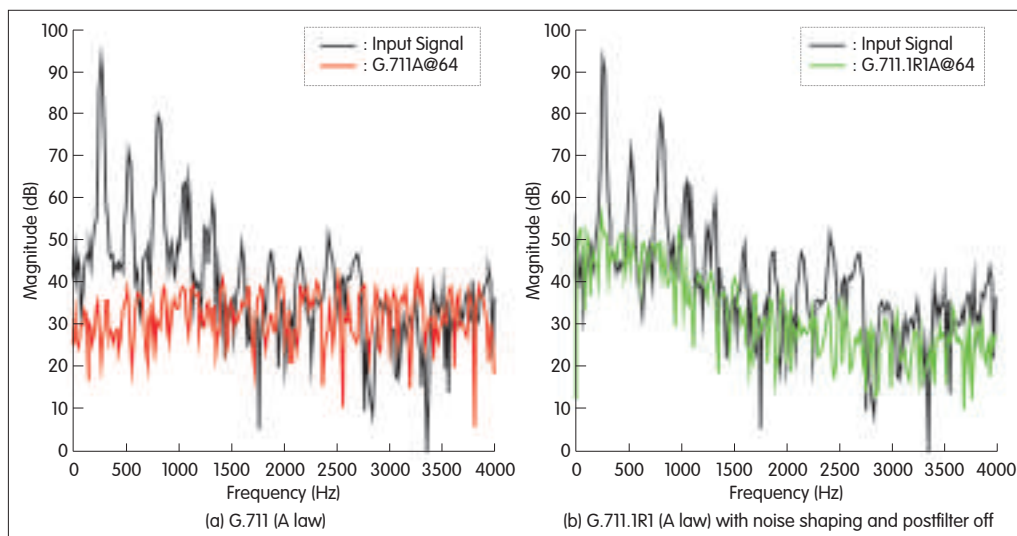
of proprietary speech codecs such as BroadVoice (BV16 and BV32) [11], [12] and SILK [13], that have structures not based on code−excited linear prediction (CELP) coding.

In this paper, we give an overview of NFC and describe a novel speech coding method that optimally combines NFC with embedded PCM or adaptive differential PCM (ADPCM). Important applications include the recent backward−compatible enhancements to G.711 and G.722.

In section 2, the principles of and approaches to noise shaping are discussed. In section 3, NFC is discussed in detail. In section 4, embedded coding and noise shaping enhancements to G.711 are described. In section 5, a counterpart to G.722 is proposed. In section 6, new codec structures inspired by NFC are discussed. Section 7 concludes the paper.

## 2 Noise Shaping Principles and Approaches

The basic idea of noise shaping is to exploit the limitations of the human auditory system, in particular, masking properties, to make coding noise inaudible. Different methods of noise shaping draw on the principles of psychoacoustics [14], [15]. In this section, we give an overview of noise shaping in speech coding but not noise shaping in perceptual audio coding. In perceptual audio coding, noise shaping is performed in the frequency or sub−band domains by a

Noise Feedback Coding Revisited: Refurbished Legacy Codecs and New Coding Models
Stéphane Ragot, Balázs Kövesi, and Alain Le Guyader



▲Figure 1. Spectrum of a narrowband female speech sample coded at 64 kbit/s and resulting noise with and without noise shaping.

masking model that determines appropriate bit allocation [16]–[18].

### 2.1 Noise Shaping in Speech Coding: G. 711

Fig. 1 shows noise shaping for a real–speech sample taken from a female who was speaking French. Two short–term spectral noise shapes are shown for PCM codecs operating at 64 kbit/s. In Fig. 1(a), 64 kbit/s PCM coding in G.711 (A law) is used [19]. The coding noise spectrum, shown in red, is nearly flat, and coding noise fills the low–energy spectral valleys of the input signal, which means the noise is clearly audible. In Fig. 1(b), a modified G.711 encoder is used that is backward–compatible with G.711 decoders. The modified G.711 encoder is defined in G.711.1 and has the same 64 kbit/s bit rate as G.711 [9], [19]. The coding noise spectrum, shown in green, approximates the envelope of the signal spectrum, which means noise is present but barely audible because it has been properly shaped.

From Fig. 1(a) and (b), a principle can be derived for perceptually optimizing speech codecs: The coding noise spectrum should approximate the signal spectrum so that a sufficient signal–to–noise ratio is maintained over each frequency range, including ranges where the signal has low energy.

### 2.2 Noise Shaping at the Encoder: Predictive Speech Coding

Noise shaping in speech coding is implemented according to the underlying coding model. For time–domain waveform coders such as PCM and ADPCM coders, noise can be shaped using NFC [3], [5]; however, related standards, such as G.711 and G.722, do not incorporate NFC [2], [20].

The most common low–bit–rate speech coders are based on linear predictive coding (LPC) (short–term prediction) and pitch (long–term prediction). To the best of our knowledge, the first temporal predictive coder that included a short–term and long–term adaptive predictor was the adaptive predictive

coder (APC) [21]. In this coder, long–term prediction is performed first. If short–term prediction was performed first, the CELP synthesis model would be obtained [22]. Atal and Schroeder recognized the necessity of shaping the quantization noise spectrum:

"The quality of the reconstructed speech can thus be improved by a suitable shaping of the spectrum of the quantizing noise so that the signal–to–noise ratio (SNR) is more or less uniform over the entire frequency range of the input speech signal." [21]

To achieve this goal, they suggested using a fixed pre–emphasis filter at the encoder side and an inverse filter at the decoder side.

Two important contributions to noise masking in speech coders were [23], [24], and these were extended by [25]–[27]. These contributions showed that noise could be shaped at the encoder by subtracting the filtered noise from the quantizer input in a feedback loop. Hence, the technique was called noise feedback coding.

The main significance of [25] was that it showed how quantization noise $Q(z)$ of an APC coder could be shaped by a prediction filter $B(z)-1$ and how a linear predictive feedback of noise could be added to the input signal to create the quantizer input that is forwarded to an APC coder with noise shaping (APC–NS) [25, Fig. 9]. The reconstructed speech $\tilde{S}(z)$ in the local decoder and in the (distant) decoder is obtained by adding the quantization noise, which is filtered by a moving–average (MA) FIR filter, to the input speech. The reconstruction speech is given by

$$\tilde{S}(z) = S(z) + B(z)Q(z) \qquad (1)$$

where $S(z)$ is the input speech, $B(z)$ is the filtered quantization noise, and $Q(z)$ is the quantization noise.

In [6] and [7], the noise masking in ADPCM coding is implemented as in [25, Fig. 9]. In [26], the quantization noise is filtered by a predictor $A(z)$, and the reconstruction noise is filtered by a predictor $B(z)$. These noises are subtracted to the input signal in the feedback loop [26, Fig. 3]. Using the notation in [26], an autoregressive moving average (ARMA) noise shape is given by

$$\tilde{S}(z) = S(z) + \frac{1-A(z)}{1-B(z)} Q(z) \qquad (2)$$

The configuration in [26, Fig. 3] leads to the same noise shaping as that in [25, Fig. 9] when $B(z)$ is set to 0. An improvement to [26, Fig. 3] is shown in [26, Fig. 7], and this improvement gives rise to a generalized predictive coder,

including a long–term predictor, with the following short–term noise–shaping characteristics:

$$\tilde{S}(z) = S(z) + \frac{1-F_s(z)}{1-P_s(z)} Q(z) \qquad (3)$$

where $P_s(z)$ is the linear predictor filter:

$$P_s(z) = \sum_{k=1}^{P} a_k z^{-k} \qquad (4)$$

and $F_s(z)$ is the weighted filter derived from $P_s(z)$ using the weighting factor $\gamma$ :

$$F_s(z) = P_s(z/\gamma) = \sum_{k=1}^{P} a_k \, \gamma^k z^{-k} \qquad (5)$$

The ARMA noise shaping of the quantization noise $Q(z)$ is usually done by the more flexible filter $H_s(z)$, given by

$$H_s(z) = \frac{A(z/\gamma_2)}{A(z/\gamma_1)} = \frac{1 + \sum_{k=1}^{P} a_k \gamma_2^k z^{-k}}{1 + \sum_{k=1}^{P} a_k \gamma_1^k z^{-k}} \qquad (6)$$

The same type of noise–shaping filter has been used in the subsequent generation of speech coders, known as analysis–by–synthesis (AbS) coders, first in multipulse coding [28] and then in CELP coding [22]. AbS is equivalent to minimizing the following CELP criterion [29]:

$$\varepsilon = \frac{1}{2\pi j} \oint_c \left| \left[ S(z) - \frac{\hat{C}(z) + g_c C_k(z)}{(1 - g_P z^{-T})\hat{A}(z)} \right] W(z) \right|^2 \frac{dz}{z} \qquad (7)$$

where $\hat{C}(z)$ is the past reconstructed excitation signal, $g_c$ is the fixed codebook gain, $C_k(z)$ is the code vector, $g_P$ is the adaptive codebook gain, $T$ is the adaptive codebook lag, $\hat{A}(z)$ is the quantized LPC filter, and $W(z)$ is the noise weighting filter (also called perceptual weighting filter). The distortion criterion $\varepsilon$ is minimized in order to whiten the weighted reconstructed noise (in square brackets, equation 7) and create a coding noise with a spectrum shape $|H_s(z)|^2 = |1/W(z)|^2$.

Fig. 2 shows the perceptual filter in narrowband predictive coding, and G.729A is used as an example.

The filter $W(z)$ in Fig. 2 is optimized so that formants get relatively lower weights; that is, more noise is tolerated in frequency ranges where the signal has more energy and can mask more noise. This masking method assumes that the masker (signal) is sufficiently well reconstructed.

Most narrowband CELP speech coders, operating at 8 kHz, use a noise–shaping filter with the form $W(z) = A(z/\gamma_1)/A(z/\gamma_2)$ or a version based on the quantized LPC filter $\hat{A}(z)$. Either $\gamma_1$ or $\gamma_2$ can be set to 0.

In G.729, the filter $W(z) = A(z/\gamma_1)/A(z/\gamma_2)$ is based on the unquantized LPC filter $A(z)$ with adaptive values for $\gamma_1$ and $\gamma_2$. In G.729A, $W(z)$ is replaced by $\hat{A}(z)/A(z/\gamma)$, where $\gamma = 0.75$ and $\hat{A}(z)$ is a quantized LPC filter used to reduce complexity. In 3GPP AMR–WB, $\gamma_2 = 0$ and the weighting filter $A(z/\gamma_1)$, where $\gamma_1 = 0.92$ in the signal domain, corresponds to the effective weighting filter used for CELP coding in a pre–emphasized signal domain $A(z/\gamma_1) / (1-\alpha z^{-1})$, where
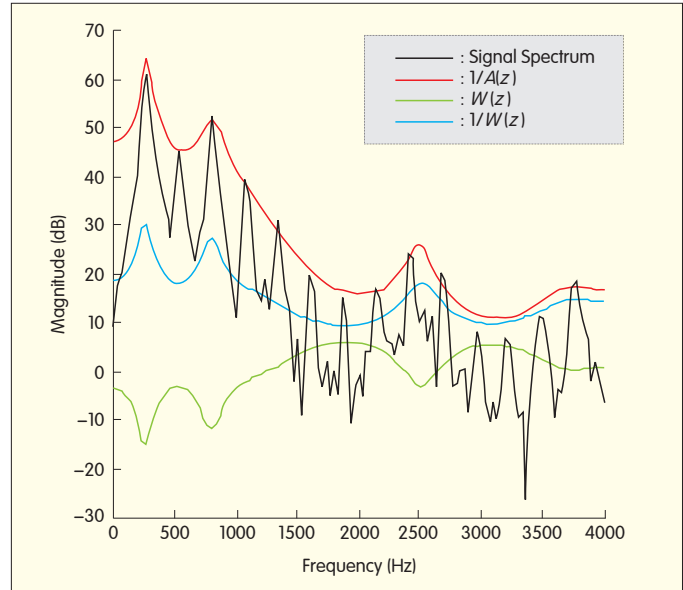


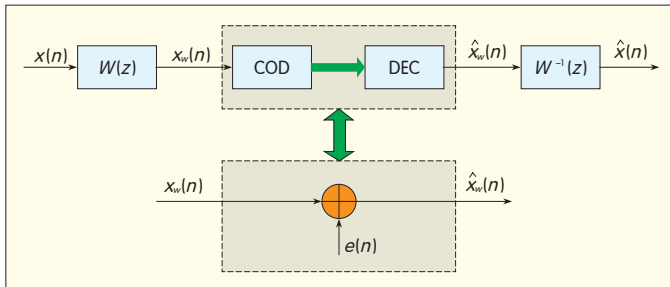▲Figure 2. Perceptual weighting filter $W(z)$ ( $\gamma$ =0.75) in G.729A. The coding noise spectrum approximates.

$\alpha$ = 0.68. The cascade of pre–emphasis and LPC filters in AMR–WB is a special case of a technique introduced in [30]. The pre–emphasis is set at $1-\alpha z^{-1}$, where $\alpha$ is not adaptive, and the weighting filter is computed only on the second filter of the cascade.

For details on predictive speech coding and AbS speech coding (including CELP) see [31]–[34]. Starting with the first CELP coding scheme in [22], the synthesis filters and perceptual weighting filters can be combined in one filter [35], which results in the filtering matrix $H$ and a matrix form of the CELP error criterion. The CELP codebook search can be further improved by backward filtering [36], using binary algebraic codes [36]–[38], or by using sparse algebraic codes that give rise to an algebraic CELP (ACELP) model [39], [31, section 17.11]. The first real–time fixed–point implementation of an algebraic CELP coder, including noise shaping, was reported in [38].

### 2.3 Noise Shaping at the Decoder Side

Noise shaping for coded speech can be performed at the decoder side using a postfilter after the speech decoder [40]. A review of postfiltering for CELP coders, including formant postfiltering, pitch postfiltering, and gain control is given in [41]. Postfiltering is based on speech–signal model parameters (LPC and pitch) that are available in typical LPC–based speech decoders but not in PCM or ADPCM coders. The underlying principle of postfiltering is to reinforce signal components and redistribute coding noise. Such filtering is described in G.729 and G.729A. In 3GPP AMR–WB, some quality–enhancement techniques are used inside the speech decoder, that is, in the LPC residual domain prior to LPC synthesis.

Noise reduction can also be used to improve quality at the decoder side. When a statistical model is available for coding

S pecial Topic

**Noise Feedback Coding Revisited: Refurbished Legacy Codecs and New Coding Models**
Stéphane Ragot, Balázs Kövesi, and Alain Le Guyader

▲Figure 3. Noise shaping by pre/postprocessing.

noise, good results can be obtained [42], [43]. Strictly speaking, this technique is not a form of noise shaping but is a reduction of the noise level by adaptive spectral attenuation, and extra delay is usually required for frequency–domain analysis and processing. If the noise is shaped at the encoder side, the resulting noise shapes need to be taken into account in decoder–side postprocessing.

### 2.4 Joint Encoder and Decoder Noise Shaping

Fig. 3 shows the principle of noise shaping by prefiltering and postfiltering. If the codec is modeled using an additive noise model, the decoded signal is

$$\hat{X}(z) = X(z) + \frac{E(z)}{W(z)}. \tag{8}$$

Noise shaping is decoupled from actual quantization or coding. The (inner) encoder–decoder can be optimized to minimize mean–square error, and the resulting coding error $E(z)$ has a nearly flat spectrum. The overall coding noise is then shaped according to the frequency response of $W^{-1}(z)$. In practice, this basic linear model is only an approximation because the filter coefficients are adapted per frame, and the actual process is non–linear.

Joint encoder/decoder processing is used in the transform coded excitation (TCX) model [44] and transform predictive coding (TPC) model [45]. The perceptual filter $W(z)$ is typically defined for LPC–based noise shaping:

$$W(z) = \frac{\hat{A}(z/\gamma_1)}{\hat{A}(z/\gamma_2)}. \tag{9}$$

The decoder needs to revert the preprocessing; therefore, the coefficients of $\hat{A}(z)$ are quantized of the linear predictive filter. This contrasts with the perceptual filters discussed in section 3, which can use unquantized coefficients.

A similar idea was explored in [46] for audio coding. In [47], an LPC filter derived from a model–based masking curve (a function of LPC and pitch) was investigated. In [48], the LPC–based perceptual weighting filter in TCX coding is modified in the frequency domain by adaptive low–frequency emphasis. This is done to improve the quality of some high–pitched music signals. In [49], combined companding and expanding at 48 kbit/s in the spectral domain results in the same quality as G.722 at 64 kbit/s [20], [50]. However, this technique implies that backward compatibility with existing coders is lost with the introduction of the new, improved

coder, and extra delay and complexity are required for frequency–domain processing.

## 3 Noise Feedback Coding

In this section, we define NFC by detailing different filter structures. We also address the related problems of noise–shaping filter estimation and loop stability. The following notation is used:

$x(n)$: input signal to (outer) encoder
$x'(n)$: input signal to (inner) encoder (modified signal including noise feedback)
$\tilde{x}(n)$: signal reconstructed by (inner) local decoder
$e(n) = \tilde{x}(n)-x(n)$: overall coding noise (or reconstruction noise)
$q(n)=x(n)-x'(n)$: inner coding noise (or quantization noise)

### 3.1 Moving–Average Structure

Fig. 4 shows noise fed back using a moving–average (MA) filter structure. The quantization noise $q(n)$ is the error introduced by the inner encoder–local decoder (COD–Local DEC) and is given by

$$q(n) = \tilde{x}(n)-x'(n). \tag{10}$$

The input signal $x(n)$ is modified by adding the quantization noise filtered by $H_A(z)-1$. The signal $X'(z)$ is given by

$$X'(z) = X(z) + (H_A(z)-1)Q(z). \tag{11}$$

Replacing $X'(z)$ with $\tilde{X}(z)$ and $Q(z)$ gives

$$\tilde{X}(z)-X(z) = H_A(z)Q(z) \tag{12}$$

Therefore, the overall reconstruction error $\tilde{X}(z)-X(z)$ is the quantization noise $q(n)$ shaped by the filter $H_A(z)$. If the bitrate is sufficient, the quantization noise can be considered white noise, and the spectrum of the reconstruction noise is close to $|H_A(z)^2|$.

The value $e_p(n)$ corresponds to a prediction because the filter $H_A(z)-1$ has a zero coefficient in $z^0$, that is, for sample $n$. This property indicates that the loop filter is an MA predictor for the next quantization error, and this prediction is based on past samples of the quantization error signal $q(n)$.

As in an APC coder, the inner coding noise $\tilde{x}(z)-x'(n)$ is the quantization noise $q(n)$. Fig. 4 can be considered a generalization of [25, Fig. 9] and [26, Fig. 3 where $A = 0$] because the noise feedback loop has been moved outside the coder. This is true when COD is a linear predictive coder without noise shaping, as in the original APC [21], PCM [2],
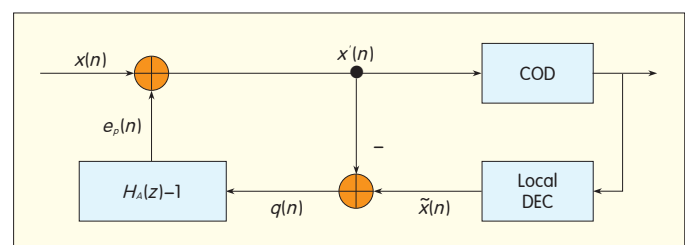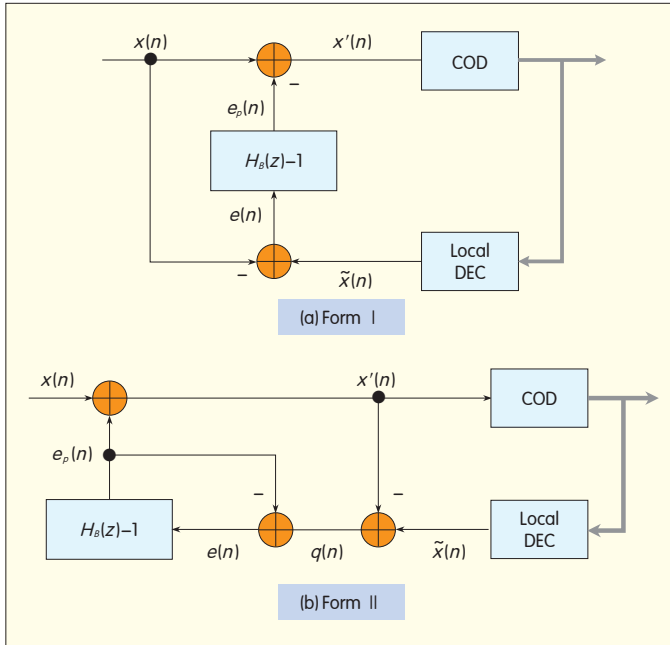


▲Figure 4. Noise feedback with moving–average structure.

**Noise Feedback Coding Revisited: Refurbished Legacy Codecs and New Coding Models**

Stéphane Ragot, Balázs Kövesi, and Alain Le Guyader



▲Figure 5. Noise feedback with auto–regressive structure.

and ADPCM [20] schemes. The structure in Fig. 4 allows the introduction of noise shaping in these legacy codecs as external pre–processing, and the COD–Local DEC loop is not affected.

### 3.2 Auto–Regressive Structure

Fig. 5 shows two equivalent auto–regressive (AR) filter structures. In form I, the decoded signal $\tilde{X}(z) = X'(z) + Q(z)$ can be expanded:

$$\tilde{X}(z) = X(z) - (H_B(z) - 1)(\tilde{X}(z) - X(z)) + Q(z) \tag{13}$$

which yields

$$\tilde{X}(z) - X(z) = \frac{1}{H_B(z)} Q(z). \tag{14}$$

Form II gives the same result. The main advantage of form I is that it saves one subtraction per sample compared with form II.

Fig. 5(a) is a special case of [26, Fig.3], where $A = 0$ and the feedback loop has again been moved outside the COD–Local DEC loop. Form II is the same as that in Fig. 4 when

$$H_A(z) = \frac{H_B(z) - 1}{H_B(z)} + 1 \tag{15}$$

Here, the loop filter is an AR predictor for the next quantization error, and the prediction is based on past samples of the coding noise $e(n)$.

### 3.3 Auto–Regressive Moving–Average Structure

Combining the structures in Figs. 4 and 5(a) gives a general auto–regressive moving–average (ARMA) structure (Fig. 6). The coding noise is given by

$$\tilde{X}(z) - X(z) = \frac{H_A(z)}{H_B(z)} Q(z) \tag{16}$$

The noise shaping in Fig. 6 can also be obtained using the structure in Figs. 4 and 5(a) by introducing an ARMA predictor in the feedback loop and computing the noise prediction from the past quantization noise and reconstructed noise, as in [51, Figs. 4 and 5]. In [51, Fig. 5] ARMA long–term noise shaping was included in the feedback loop to improve the performance of legacy PCM and ADPCM coders for voice signals.

### 3.4 Loop Filter Design: HA and HB

The simplest design for the loop filter includes a fixed filter, such as $H_A(z) = 1 + z^{-1}$. In this case, the effective quantization noise is colored in a predefined way. For speech or audio signals, this approach is generally not optimal, and it is better to adapt the loop filter to the short–term spectral properties of the input signal [5]. A detailed analysis using high–rate distortion theory (i.e. small errors) shows that, with certain assumptions, the optimal loop filter whitens the input signal [5]. Hence, a linear predictive (LP) analysis can be used for loop–filter estimation.

In [25] and [26], LPC techniques were used to obtain the weighting filters from the linear predictive coefficient. In [52], the weighting filter is an LPC filter computed on the adaptively pre–emphasized speech signal.

Because noise feedback is implemented at the encoder side, the loop filter estimation can be based on the original signal $x(n)$. In cases such as embedded PCM, where a core coder and enhancement stages depend on the loop filter at both the encoder and decoder sides, it can be beneficial to estimate the filter on the decoded core signal $\tilde{x}(n)$. This is the case of G.711.1, where the loop filter is estimated through an LP analysis on the decoded core signal $\hat{S}_{Lo}(n)$ at the encoder and decoder. In G.722, the LP analysis is done on the input signal $x(n)$.

### 3.5 Loop Stability

As in any system with feedback, the noise feedback loop may become unstable in certain conditions. The quantization process is non–linear, and this makes it difficult to model the
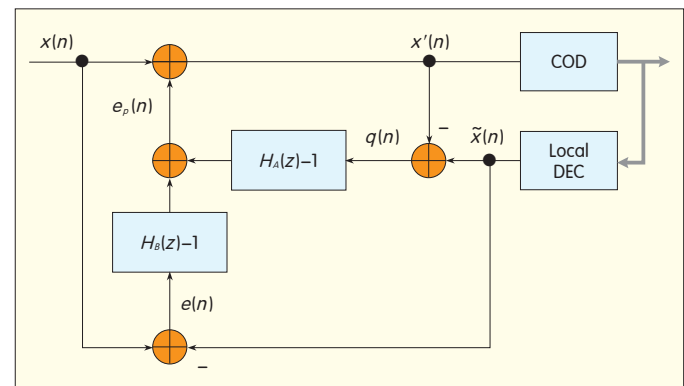


▲Figure 6. Noise feedback with ARMA structure.
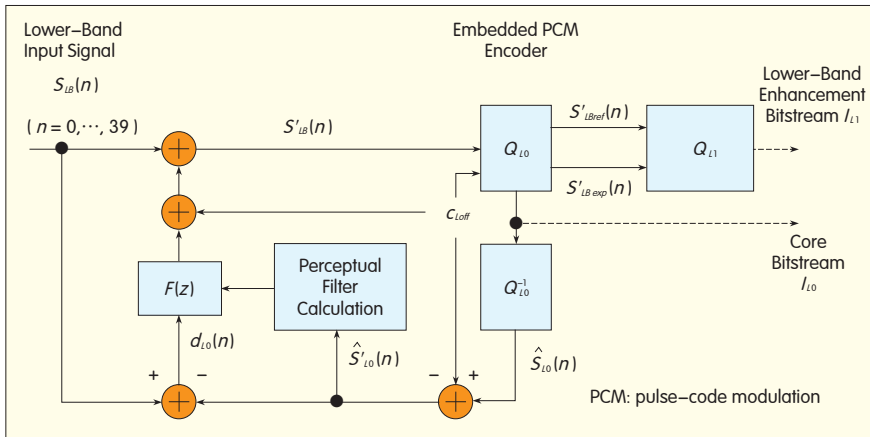
▲Figure 7. A G.711.1 encoder [58].

loop behavior. For the loop transfer function to be stable, the NFC transfer function must have its poles inside the unit circle. A stability analysis of the APC coder with a parametric model of the APC system can be found in [53].

In [54], stability was analyzed for the embedded ADPCM with noise feedback. Perceptual input/output SNR of the coder is given by

$$SNR_P = G_P \left[ \frac{SNR_Q - 1}{E_D} + 1 \right] \qquad (17)$$

where $G_P$ is the ADPCM prediction gain, $SNR_Q$ is the SNR of the quantizer (about 24 dB for a 5-bit quantizer), and $E_D$ is the impulse response energy of the masking filter $H_s$ (in the general case). When $SNR_P$ is low, the ADPCM coder is close to being unstable.

Because of this problem, the order of the loop filter is usually small. In recent G.711.1 and G.722B standardization work, an AR structure is used with a loop filter of order 4, and some heuristic methods have been developed that limit loop instability by detecting specific signals, for example, chirps and combinations of sinusoids and frequency hops. These signals create instability or degrade performance compared with having no feedback. Stability issues are mitigated by detecting such signals, and the noise-shaping filter resonances are attenuated or the loop is turned off for a period of time [9], [10], [54].

# 4 Improving G. 711 by Using Embedded PCM Coding with Noise Shaping

G.711 [2] is the most widely deployed speech codec in fixed networks and voice-over-IP (VoIP) networks. In [2], the input linear PCM is coded at 8 bits/sample using either A-law or a μ-law. In 2007, ITU-T SG16 started a work item called G.711-WB (wideband) under the initiative of NTT. The objective of this work was to develop a 64-80-96 kbit/s embedded extension of G.711 that was low-delay, low-complexity, and capable of 50-7000 Hz wideband [55]. The principles of embedded PCM, ADPCM, CELP coders, and hierarchical coders such as G.729.1 [56] can be found

in [57].

## 4.1 Principle of G. 711.1

The G.711.1 encoder receives 8 kHz or 16 kHz input signals. The input is divided into 5 ms frames. The encoding bit rate can be set at 64 kbit/s, 80 kbit/s, or 96 kbit/s. The input signal is decomposed into two sub-bands using a quadrature mirror filter bank (QMF). The 0-4000 Hz lower band is coded by an embedded PCM encoder that is interoperable with G.711 (Fig. 7). The 4000-8000 Hz higher band is coded in the modified discrete cosine transform (MDCT) domain.

The bitstream structure of G.711.1 is defined in Table 1. More details on G.711.1 can be found in [9], [58].

## 4.2 Embedded PCM Coding Without Noise Shaping

The PCM coding in [2] is similar to the scientific notation; a given number of bits (resolution) in linear PCM is retained to form the mantissa, and an exponent value gives the correct scaling. In G.711, which has 8 bits per sample, 5-bit precision is used for the signed mantissa, and this results in approximately 38 dB SNR (except for the smallest values in the first segment). The most significant bit (MSB) is always 1 in natural binary decomposition; therefore, it does not need to be transmitted. Only the next 4 bits (after MSB) are sent to the decoder; one bit is reserved for the sign, and the remaining 3 bits are used to encode 8 possible exponent values (segment indices).

In G.711.1, a 16 kbit/s, 2 bit/sample enhancement layer was introduced to enhance the 0-4000 Hz lower band. In [55], two extra mantissa bits of the binary representation of input samples are extracted and transmitted to the extension layer, which gives an overall resolution of 6 bits in the first segment and 7 bits in the other segments. In the final tuning phase of G.711.1 standardization, a dynamic bit allocation was taken to allocate 1, 2 or 3 bits per sample to the frame bit budget [9].

## 4.3 Embedded PCM Coding with Noise Shaping

In the 0-4000 Hz lower band of G.711.1, noise shaping was introduced to PCM coding to improve the quality of the interoperable core bitrate mode at 64 kbit/s and to improve the quality of the enhanced high-bitrate mode at 80 kbit/s. In

▼Table 1. Bitstream structure of G.711.1

| Mode | Sampling Rate (kHz) | Core Layer $I_{L0}$ | Lower-Band Enhancement Layer $I_{L1}$ | Higher-Band Enhancement Layer $I_{L2}$ | Overall Bit Rate (kbit/s) |
| --- | --- | --- | --- | --- | --- |
| | | 64 kbit/s | 16 kbit/s | 16 kbit/s | |
| R1 | 8 | x | − | − | 64 |
| R2a | 8 | x | x | − | 80 |
| R2b | 16 | x | − | x | 80 |
| R3 | 16 | x | x | x | 96 |

S pecial Topic

Noise Feedback Coding Revisited: Refurbished Legacy Codecs and New Coding Models
Stéphane Ragot, Balázs Kövesi, and Alain Le Guyader

this way, the 64 kbit/s core layer of G.711.1 can be decoded by a legacy G.711 decoder, and the quality of G.711 encoding and decoding can be significantly improved, especially for low–level clean speech.

In the G.711.1 encoder, noise shaping is performed on the core–layer signal using the form I AR structure (Figs. 5(a) and 7). The decoded core coder output is given by

$$\hat{S}_{L0}(z) = S_{LB}(z) + \frac{1}{1+F(z)} Q(z) \qquad (18)$$

The enhancement layer is defined in section 4.1, and noise shaping had to be performed on the enhancement signal at the decoder before being added to the core decoded signal. To prevent a mismatch between the decoder and the encoder, the perceptual filter $F(z)$ in Fig. 7 is calculated using the past decoded core signal available at both encoder and decoder side [51], [52].

In the encoder, the quantization noise of the core layer filtered by $F(z)$ is fed back to be added to the input signal encoded using embedded PCM coding (Figs. 5a and 7). Therefore, if the decoder decodes only the core layer, the quantization noise is already shaped when it is decoded by a G.711 decoder. If the enhancement layer is also decoded, the enhancement–layer contribution, that is, the difference between the core bitrate output and the higher bitrate output, has to be shaped at the decoder. It is thus shaped by the noise shaping filter $1/(1+F(z))$ prior to being added to the core layer output. Noise shaping is performed at the encoder side for the core layer and at the decoder side for the enhancement layer [51], [52].

In the G.711.1 core layer, the ordinary log–PCM encoding is replaced by a dead–zone quantizer encoding scheme [9, clause 7.3.3]. Another way to enhance the quality of the PCM coding and reduce the perceptual impact of PCM quantization noise is to postfilter quantization noise at the decoder [42]. This technique is detailed in [9, Appendix I].

### 4.4 G.711 App. III: Noise Shaping as Preprocessing to G. 711 and Other Tools

Noise shaping was incorporated into the 64 kbit/s core coder in G.711.1; however, it can also be performed externally as preprocessing to G.711. This preprocessing approach, together with other tools such as the postfiltering of G.711.1 App. I, form G.711 Appendix III [19].

## 5 Improving G.722 by Using Embedded ADPCM Coding with Noise Shaping

G.722 was the first normalized 50–7000 Hz wideband (WB) speech codec [20]. This coder is based on a sub–band embedded ADPCM coding scheme [50]. The input signal of G.722, sampled at 16 kHz, is decomposed by a QMF into two sub–bands: 0–4000 Hz and 4000–8000 Hz. Each sub–band is coded separately using ADPCM coding. The total bit rate is 64, 56 or 48 kbit/s, depending on the number of bits allocated to the lower band. The block diagram of the G.722 encoder and decoder can be found in [50, Figs. 9 and 12]. G.722 is the
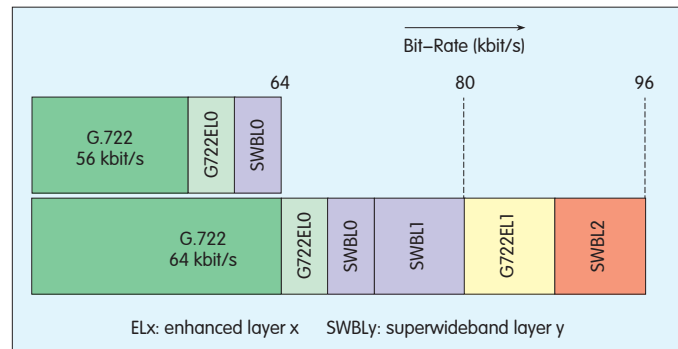


▲Figure 8. Bitstream structure of G.722B.

wideband codec specified for DECT new Generation (DECT–NG) terminals, which are also called Cordless Advanced Technology—Internet and quality terminals. This codec provides better audio quality than the legacy narrowband DECT codec (G.726). To further improve call quality, 50–14,000 Hz superwideband (SWB) coding is the next step. In 2008, the ITU–T launched the G.711.1/G.722 SWB work item under the initiative of France Telecom and NTT. This activity resulted in G.722 Annex B (G.722B) [10], [59]. We focus here on improvements made to G.722 as part of G.722B development. The principle of embedded coding is explained in [57].

### 5.1 Principle of G. 722 Annex B

G.722–SWB standardization is aimed at developing an embedded scalable SWB extension of G.722 operating at 64, 80, and 96 kbit/s. The 64 kbit/s bit rate of G.722B was specifically designed to fit into existing 64 kbit/s CAT–iq transport channels which comprise wideband G.722 at 56 kbit/s and 8 kbit/s for an SWB enhancement layer. Fig. 8 shows the embedded bitstream structure of G.722B. Here we only discuss the part of G.722B that corresponds to the G.722 core bitstream layer at 56 kbit/s and 64 kbit/s.

The main objective of G.722B is to extend G.722 to SWB. As with G.711.1, it was necessary to improve G.722 because artifacts in the lower band are reinforced when higher–band extensions are added. At 64 kbit/s, G.722B operates with a 56 kbit/s G.722 core coder that has an 8 kbit/s enhancement layer split into a G.722 high–band enhancement (EL0) and an SWB enhancement (SWBL0). At 80 or 96 kbit/s, G.722B operates with a 64 kbit/s G.722 core coder that has two high–band enhancement layers: G722EL0 and G722EL1, and three SWB enhancement layers: SWBL0, SWBL1, and SWBL2.

As shown in Fig. 9, the G.722B coder includes an enhanced G.722 coder for the wideband (WB) part, which enhances the G.722 4000–8000 Hz higher band, and a bandwidth extension in the MDCT domain.

In the original G.722 coder [20], [50], the lower sub–band is coded by an embedded ADPCM coder at 6, 5 or 4 bit/sample, and the higher sub–band is coded by an ADPCM coder at 2 bit/sample. G.722 relies on plain ADPCM coding with no noise shaping. In the following, we describe the G.722 lower–band (LB) coder and the G.722 higher–band (HB)
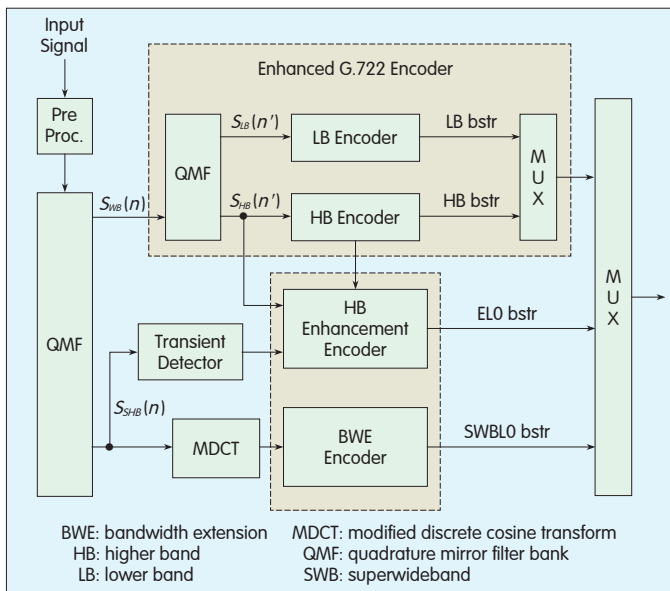
▲Figure 9. Block diagram of G.722B.
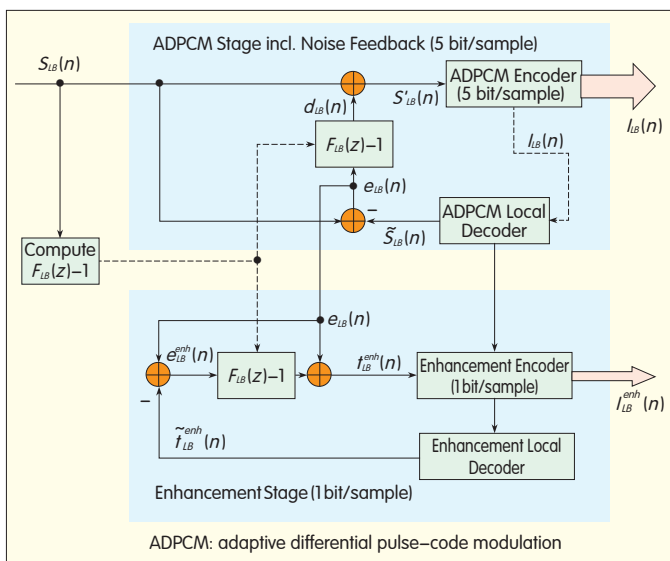


▲Figure 10. Lower band coding in G.722B.

coder, including noise shaping.

### 5.2 Improved 0–4000 Hz Lower–Band Embedded ADPCM Coding Including Noise Feedback

Fig. 10 shows a modified LB ADPCM coder, including noise shaping. In the case of G.722B at 64 kbit/s, the G.722 LB encoder operates at 5 bit/sample, and in the case of G.722B at 80 or 96 kbit/s, the G.722 LB encoder operates at 6 bit/sample. To shape the coding noise at 5 and 6 bit/sample in a coherent and optimal way, embedded ADPCM coding is modified so that it operates in two optimal stages. A noise feedback loop is included in the first stage, and AbS is done in the second stage. Therefore, a noise feedback loop is used with the ADPCM coder operating at a reduced bitrate of

40 kbit/s (5 bit/sample). In Fig. 10, noise shaping is performed on the core layer signal using form I of the AR structure shown in Fig. 5(a). The COD is an ADPCM coder with 5 bit/sample, and the Local DEC is an ADPCM local decoder with 5 bit/sample. An enhancement encoder using AbS and operating at 1 bit/sample brings the bit rate to 48 kbit/s (6 bits/sample) [10], [60].

Thus, the modified G.722 LB encoder relies on embedded scalar quantization, which is done by AbS in a perceptually weighted domain. Both stages use the same noise shaping filter $F_{LB}(z)-1$ derived from a linear predictive coding (LPC) filter $F_{LB}(z)$ with an order of 4.

### 5.3 Improved 4000–8000 Hz Higher–Band ADPCM Coding (G.722 HB and EL0/EL1)

Fig. 11 shows the HB encoder. The core G.722 HB encoder corresponds to the G.722 HB ADPCM encoder operating at 16 kbit/s (2 bit/sample). The ADPCM signal–to–noise ratio in the higher band is low because the allocated bitrate is 2 bit/sample as opposed to 5 or 6 bit/sample for G.722 LB ADPCM coding. This difference in bitrate means that contrary to the G.722 lower band, no noise feedback is used in higher–band ADPCM coding. For noise feedback to be efficient, a bit rate higher than 2 bit/sample is desirable. If additional bit rate is available to improve the quality of the G.722, these extra bits are allocated to the higher band. The quantization resolution can be improved by the enhanced G.722 HB encoder, which relies on embedded scalar quantizers of 0.475 bit/sample (EL0) and 1 bit/sample (EL1).

The G.722 HB ADPCM encoder can be extended by a scalable or embedded bitstream. This extension is done in two embedded stages using AbS, which is similar to the approach used for low–band enhancement. The first extension stage, G722EL0, is performed at 3.8 kbit/s, and 19 of the 40 samples are used only in non–transient frames to enhance quality with 1 bit/sample. This forms the G722EL0
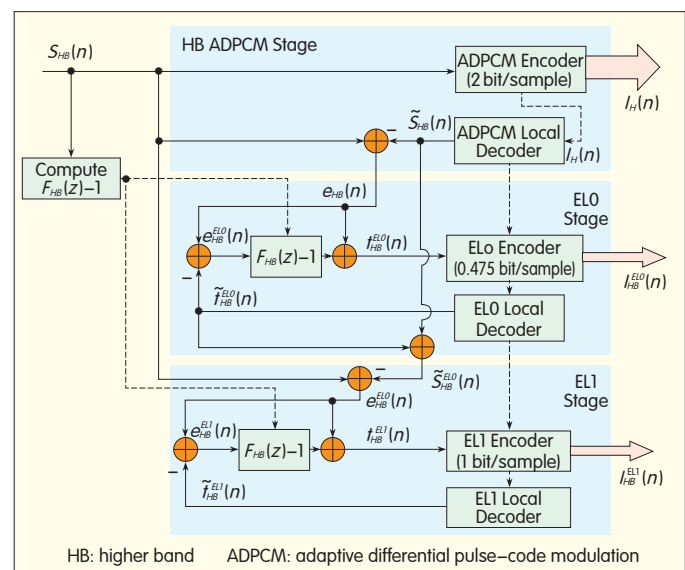


▲Figure 11. Higher band coding in G.722B.

layer, which is used in all superwideband layers (Fig. 8). The G722EL0 layer is disabled in the case of transient signal segments, where the spare 19 bits are allocated to SWB extension. The second extension stage, G722EL1, is performed at 8 kbit/s (1 bit/sample) to further refine the quantization of the G.722 higher band, and this forms the G722EL1 layer that is only used in G.722B at 96 kbit/s.

### 5.4 Analyzing the Core Coder Enhancement Layer with Analysis–by–Synthesis in Embedded ADPCM Coding

Here, we explain how the LB embedded ADPCM coder in G.722 Annex B [10] is modified to allow backward–compatible noise shaping. As described in Section V.B of the G.722 standard, the G.722 LD encoder is split into a core coder with noise shaping at 56 kbit/s (5 bit/sample) and an AbS enhancement to reach 6 bit/sample.

The legacy ADPCM local decoder at 5 bit/sample reconstructs the output signal by adding the scaled ADPCM quantizer output at 5 bit/sample to the prediction obtained at 4 bit/sample. To obtain the 6th bit in the legacy coder (without noise shaping), the embeddedness of the ADPCM quantizer can be used to find the two scaled and adequate 6–bit quantizer levels. The predicted signal can be added to the quantizer output to give two possible synthesized signals, and then a search can be done to determine which of the two signals is the closest to the input signal $s_{LB}(n)$. AbS is done to find the extension layers of the core coder bit by bit, and it does not involve any noise shaping.

To introduce a weighted–error criterion [60], the extension–stage error criterion is

$$E_{LB} = \frac{1}{2\pi j} \oint_c \left| \left\{ [S_{LB}(z) - \tilde{S}_{LB}(z)] - \tilde{\xi}^i_{LB}(z) \right\} W(z) \right|^2 \frac{dz}{z}, \ i = 0, 1 \quad (19)$$

where $\tilde{S}_{LB}(z)$ is the core coder reconstructed signal, $\tilde{\xi}_{LB}(z)$ is the z–transform of the enhancement output signal, and $W(z)$ is the weighting filter equal to $F_{LB}(z) = 1 + \sum_{i=1}^{4} a_{LBi} z^{-i}$.

The signal $\tilde{\xi}^i_{LB}(z)$ is equal to the past outputs of the enhancement layer for $n' < n$, that is, $\tilde{t}^{enh}_{LB}(n)$ in Fig. 10, and is equal to the ith enhancement candidate $\xi^i_{LB}(n')$ for $n' = n$ [10, B.6–32]. The noise spectrum of $S_{LB}(z) - \tilde{S}_{LB}(z) - \tilde{\xi}^{i_{opt}}_{LB}(z)$ at the output of the enhancement stage has the shape $\left| \frac{1}{F_{LB}(z)} \right|^2$.

The signal $S_{LB}(z) - \tilde{S}_{LB}(z)$ in the bracket in the error criterion is equal to $e_{LB}(n)$ in Fig. 10.

Although the coding structure in Figs. 10 and 11 use AbS, this technique is quite different from CELP coding because the sample–by–sample gain is provided by the core layer. Also, for each sample $n$, the enhancement codebook is obtained from the ADPCM index of the core layer and the possible preceding enhancement layers. It is obtained using the embeddedness of the ADPCM quantizers [10, B.6–32].

The same method can be used for the high–band ADPCM encoding. Because the method is based on AbS with noise shaping, the results can be extended to minimize error of (Eq. 19) on a block basis and to increase the efficiency of the noise shaping. If care is taken in embedded quantization, the

reconstructed, enhanced codebook with two entries $\xi^i_{LB}(n)$ $i = 0, 1$ (19) can be extended to a binary algebraic tree codebook, which allows fast algorithms to be derived for minimizing the error criterion [61]. Computation load is only slightly increased. Therefore, it is possible to further improve the quality of the G722EL0 and G722EL1 layers.

### 5.5 Experimental Results

In the official ITU–T tests for G.722B, only SWB cases were tested. In [62], some additional wideband tests were done to assess the backward compatibility of G.722B with G.722.

## 6 New Codec Structures

The principles of noise feedback were recently revisited in an effort to develop new codec structures, for example, BV16, BV32 [11], [12] and SILK [13].

### 6.1 BV16 and BV32 Coders

The BV16 and BV32 coders [11], [12] are based on a synthesis model that includes a long–term (LTP) synthesis filter and a short–term (LPC) synthesis filter. The first predictive speech codecs that used noise shaping [25], [26] relied on a scalar quantization and a short–term noise–feedback filter to shape the spectral envelope of the coding noise. In contrast, the two–stage noise feedback coding (TSNFC) in [11] and [12] relies on two NFC stages in a nested loop. In the first NFC stage, short–term prediction and short–term noise spectral shaping (spectral envelope shaping) is performed. In the second (nested) NFC stage, long–term prediction and long–term noise spectral shaping (harmonic shaping) is performed.

In Fig. 12, the long–term noise shaping predictor $N_l(z) - 1$ is located around the quantizer, as in [25, Fig. 7]. This implies MA long–term noise shaping by $N_l(z)$. The short–term noise–shaping predictor $F_s(z)$ acts on the noise of the long–term quantizer loop so that the noise is shaped by $1 - F_s(z)$. Thus, the short term filter $1 - P_s(z)$ is a pre–emphasis, and the coding noise is shaped by $\frac{1}{1 - P_s(z)}$ [21]. To more precisely describe the z–transform of the overall system, the output speech is given in terms of the input:

$$S_Q(z) = S(z) - N_l(z) \frac{1 - F_s(z)}{1 - P_s(z)} \quad (19)$$

MA long–term noise shaping and ARMA short–term noise shaping depend on the LPC predictor $P_s(z)$, so the amount of noise shaping is obtained from the damping factor $\gamma$ used in $F_s(z)$. An alternative ARMA long–term noise shaping and more flexible ARMA short–term noise shaping is described in [51, Fig. 5]. Further details on the BV16 or BV32 coders can be found in [11], [12] and [31].

### 6.2 SILK Coder

SILK is a speech codec designed for VoIP [13]. The synthesis model for a SILK coder is based on a long–term synthesis filter and a short–term synthesis filter. The excitation
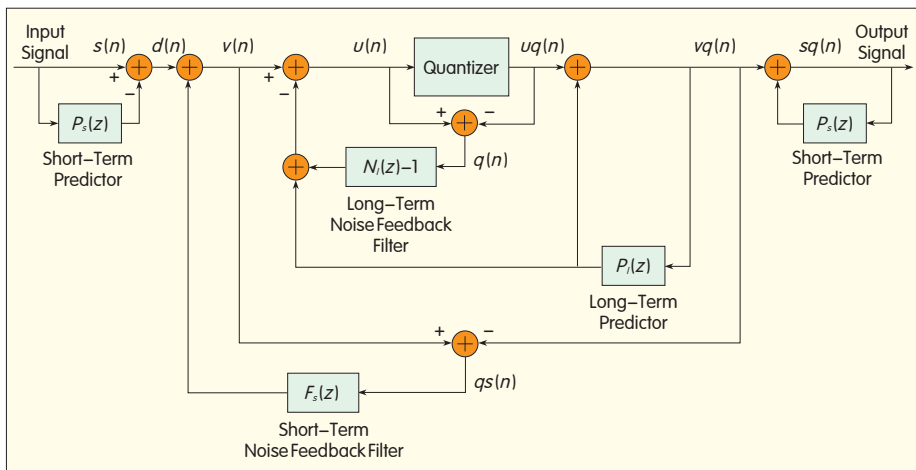
▲Figure 12. Noise shaping in BV32 coder.
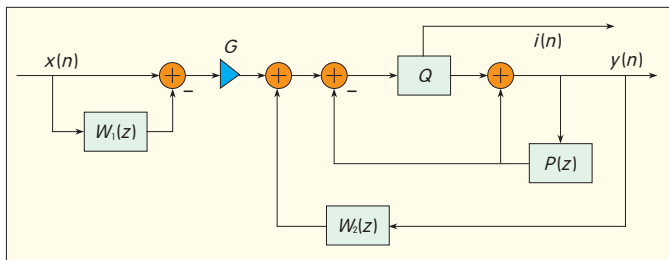


▲Figure 13. SILK coder.

signal is obtained by range−coded quantization, which is similar to arithmetic−coded quantization. Fig. 13 shows the noise shaping of the SILK codec.

The noise−shaping predictor $W_2(z)$ is in a feedback path around an inner coder. This inner coder has a generalized predictor filter $P(z)$ that includes LPC and LTP. Therefore, the quantization noise $Q(z)$ is filtered by $\dfrac{1}{1-W_2(z)}$ . The filter $1-W_1(z)$ acts as a pre−filter whose inverse is not present in the local decoder, as in the classical pre−filter in [21]. As a consequence, the spectral content of the input speech is modified by the codec. This can be confirmed by writing the input/output $z$−transform of Fig. 13 as

$$Y(z) = G \frac{1-W_1(z)}{1-W_2(z)} X(z) + \frac{1}{1-W_2(z)} Q(z) \qquad (20)$$

If $G$ is set at 1 and $W_1(z) = W_2(z)$, the coder of Fig. 5(a) is obtained. Noise shaping in SILK involves modifying the input signal. This is similar to postfiltering the input as in [41], where the filter has a short−term and long−term section and gain is controlled by $G$, and shaping the noise, as in NFC.

The SILK noise shaping structure can be obtained by transforming the structure in Fig. 5(a). In the structure in Fig. 5(a), noise is shaped in the vector case by minimizing the weighted error criterion between the input and output signals. The structure in Fig. 13 allows the levels of the decoded speech formants to be matched with the levels of the original speech formants. The levels of the spectral valleys are

decreased relative to the levels of the spectral peaks, including speech formants and harmonics. To obtain both long−term and short−term noise shaping, $W_1(z)$ and $W_2(z)$ should be of the form

$$W_{i=1,2}(z) = [Sh_{i=1,2}(z) \\ + LT_{i=1,2}(z)[1-Sh_{i=1,2}(z)]] \qquad (21)$$

where $SH_{i=1,2}(z)$ is the short−term predictor and $LT_{i=1,2}(z)$ is the long−term predictor. Then $W_{i=1,2}(z)$ appears to be the product of the short−term and long−term inverse filters minus one.

## 7 Conclusion

Two applications of NFC were described: the improvement of existing coders, for example, G.711 and G.722 by making them backward−compatible, and the development of new codec structures that are alternatives to the ubiquitous CELP coding model.

The improvement of legacy codecs is an important research topic, and significant gains are possible with refurbished codecs. Versions of G.711 and G.722 using both encoder−side and decoder−side techniques have been standardized by the ITU−T [9], [10], [19].

As part of enhanced voice standardization (EVS), 3GPP SA4 is also considering improving the 3GPP AMR−WB standard by making it bit−stream interoperable.

**References**

[1] R. M. Dolby, "An audio noise Reduction System," *JAES*, vol. 15, no. 4, pp. 383−388, Oct. 1967.
[2] *Pulse code modulation (PCM) of voice frequencies*, ITU−T G.711, Dec. 1972.
[3] C. C. Cutler, "Transmission systems employing quantization," U.S. Patent 2 927 962, 1960.
[4] H. A. Spang and P. M. Schultheiss, "Reduction of quantizing noise by use of feedback," *IRE Trans. Communications Systems*, vol. 10, issue 4, pp. 373−380, Dec. 1982.
[5] N. S. Jayant and P. Noll, "Noise feedback coding," Chapter 7 in *Digital Coding of Waveforms*, New Jersey: Prentice Hall, Mar. 1984.
[6] H. S. Lee and C. K. Un, "On the performance of speech waveform coder with noise spectral shaping", in *IEEE Trans. Commun.*, vol. 33, no.7, pp. 742−746, July 1985
[7] H. D. Kim and C.K. Un, "Embedded ADPCM with noise shaping for packet voice transmission," in *IEEE Electronic Letters*, vol. 3, no. 5, Feb. 1987
[8] H. Inose, Y. Yasuda, and J. Marakami, "A telemetering system by code modulation: modulation," *IRE Transations on Space Electronics Telemetry*, vol. SET−8, pp. 204−209, Sept. 1962.
[9] *Wideband embedded extension for G.711 pulse code modulation*, ITU−T G.711.1, March 2008.
[10] N*ew Annex B with superwideband embedded extension*, ITU−T G.722 Amd. 1, Nov. 2010.
[11] J. H. Chen, "Novel codec structures for noise feedback coding of speech," in Proc. *Int. Conf. Acoustic, Speech and Signal Processing (ICASSP)*, Toulouse, 2006, vol. 1, pp. 1.
[12] J. H. Chen and J. Thyssen, "The broadvoice speech coding algorithm," in Proc. *Int. Conf. Acoustic, Speech and Signal Processing (ICASSP)*, Honolulu, HI, 2007, vol. 4, pp. 537−540.
[13] K. Vos, S. Jensen, and K. Soerensen. (2010, Sept.). *SILK speech codec* [Online].

# S pecial Topic

**Noise Feedback Coding Revisited: Refurbished Legacy Codecs and New Coding Models**

Stéphane Ragot, Balázs Kövesi, and Alain Le Guyader

Available: http://tools.ietf.org/html/draft–vos–silk–02

[14] E. Zwicker and H. Fastl, *Psychoacoustics: Facts and Models*, 2nd edition, Springer, 1999.

[15] W. H. Hartmann, *Signals, Sounds, Sensation*, New York: Springer, 1997.

[16] K. Brandenburg, "Perceptual coding of high quality digital audio," Chapter 2 in *Applications of Digital Signal Processing to Audio and Acoustics*, Dordrecht: Kluwer Academic Publishers, 2002.

[17] M. Bosi and R. E. Goldberg, *Introduction to Digital Audio Coding and Standards*, New York: Springer, 2002.

[18] J. Herre and M. Lutzky, "Perceptual audio coding of speech signals", Chapter 18 in *Springer Handbook of Speech Processing*, (Benesty, Sondhi, Huang Editors), New York: Springer, 2008.

[19] *Audio quality enhancement toolbox*, ITU–T G.711 App. III, Nov. 2009.

[20] *7 kHz audio–coding within 64 kbit/s*, ITU–T G.722, Nov. 1988.

[21] B. S. Atal and M. R. Schroeder, "Adaptive predictive coding of speech signals," *Bell System Tech. J.*, vol. 49, no. 8, pp.1973–1986, Oct. 1970.

[22] M. R. Schroeder and B. S. Atal, "Coded–excited linear prediction (CELP): high quality speech at low bit rates," in Proc. *Int. Conf. Acoustic, Speech and Signal Processing (ICASSP)*, Tampa, FL, vol. 10, pp. 937–940, 1985.

[23] M. Berouti and J. Makhoul, "High quality adaptive predictive coding of speech," in Proc. *Int. Conf. Acoustic, Speech and Signal Processing (ICASSP)*, Tulsa, AZ, pp. 303–306, Apr. 1978.

[24] B. S. Atal and M. R. Schroeder, "Predictive coding of speech signals and subjective error criteria," in Proc. *Int. Conf. Acoustic, Speech and Signal Processing (ICASSP)*, Tulsa, AZ, pp. 573–576, Apr. 1978.

[25] J. Makhoul and M. Berouti, "Adaptive noise spectral shaping and entropy coding in predictive coding of speech," *IEEE Trans. Acoustics, Speech and Signal Processing*, vol.27, no. 1, pp. 63–73, Feb. 1979.

[26] B. S. Atal and M. R Schroeder, "Predictive coding of speech signals and subjective error criteria," *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. 27, no. 3, pp. 573–576, June 1979.

[27] M. R. Schroeder, B. S. Atal, and J. L. Hall, "Optimizing digital speech coders by exploiting masking properties of the human ear," *J. Acoust. Soc. Amer.*, vol. 66, issue 6, pp. 1647–1652, 1979.

[28] B. S. Atal and J. R. Remde, "A new model of the LPC excitation for producing natural–sounding speech at low bit rates," in Proc. *Int. Conf. Acoustic, Speech and Signal Processing (ICASSP)*, Paris, pp. 614–617, 1982.

[29] A. Le Guyader, R. Di Francesco, and C. Lamblin, "Derivation of efficient CELP coding algorithms using the Z–transform approach," *Proc. ICASSP*, Toronto, vol. 1, pp. 209–212, 1991.

[30] C. Quinquis, A. Le Guyader, "Method of analysing by linear prediction an audio frequency signal, and its application to a method of coding and decoding an audio frequency signal," Patent EP0782128, Jul. 2, 1997.

[31] J. H. Chen and J. Thyssen, "Analysis–by–synthesis speech coding", Chapter 17 in *Springer Handbook of Speech Processing* (Eds. Benesty, Sondhi, Huang), Springer, 2008.

[32] A. M. Kondoz, *Digital Speech Coding for Low Bit Rate Communication Systems*, New York: Wiley, 2004.

[33] P. Vary and R. Martin, *Digital Speech Transmission – Enhancement, Coding & Error Concealment*, New York: Wiley, 2006.

[34] N. Moreau, *Tools for Signal Compression: Applications to Speech and Audio Coding*, New York: Wiley, 2011.

[35] I. M. Trancoso and B. S. Atal, "Efficient procedures for finding the optimum innovation in stochastic coders," in Proc. *Int. Conf. Acoustic, Speech and Signal Processing (ICASSP)*, Tokyo, vol. 11, pp. 2375–2378, 1986.

[36] J. P. Adoul, P. Mabilleau, M. Delprat, and S. Morisette, "Fast CELP coding based on algebraic codes," in Proc. *Int. Conf. Acoustic, Speech and Signal Processing (ICASSP)*, Dallas, TX, pp. 1957–1960, 1987.

[37] D. Massaloux, A. Le Guyader, and J. F. Zurcher, "A new fast algorithm used in a vector adaptive predictive coder," CNETreport, ref:296/LAA/TSS/CMC, 1986.

[38] A. Le Guyader, D. Massaloux, and J. F. Zurcher, "A robust and fast CELP coder at 16 kbit/s," *Speech Communication*, vol.7, no. 2, pp. 217–226, Jul. 1988.

[39] R. Salami and al., "Design and description of CS–ACELP: a toll quality 8 kb/s speech coder," *IEEE Trans. Speech and Audio*, vol. 6, no. 2, pp. 116–130, Mar. 1998.

[40] V. Ramamoorthy and N. Jayant, "Enhancement of ADPCM speech by adaptive post–filtering," *AT&T Bell Laboratories Tech. J.*, vol. 63, no. 8, pp. 1465–1475, Oct. 1984.

[41] J. H. Chen and A. Gersho, "Adaptive postfiltering for quality enhancement of coded speech," *IEEE Trans. on Speech and Audio Processing*, vol. 3, no. 1, pp. 59–71, Jan. 1995.

[42] J. Garcia, C. Marro, and B. Kövesi, "A PCM coding noise reduction for ITU–T G.711.1," in Proc. *Interspeech*, Brisbane, Australia, pp. 57–60, Sept. 2008.

[43] C. M. Konate, "Enhancing speech coder quality: improved noise estimation for postfilters," M.Eng. thesis, Dept. Elec. & Comp. Eng., McGill University, June 2011.

[44] R. Lefebvre, R. Salami, C. Laflamme, and J. P. Adoul, "High quality coding of wideband audio signals using transform coded excitation (TCX)," in Proc. *ICASSP*, Adelaide, vol. 1, pp. 193–196, 1994.

[45] J. H. Chen and D. Wang, "Transform predictive coding of wideband speech signals," in Proc. *Int. Conf. Acoustic, Speech and Signal Processing (ICASSP)*, Atlanta, GA, vol. 1, pp. 275–278, 1996.

[46] B. Edler and G. Schuller, "Audio coding using a psychoacoustic pre– and post–filter," Proc. *ICASSP*, Istanbul, vol. 2, pp. 881–884, 2000.

[47] S. Bruhn, V. Grancharov, B. Kleijn, J. Klejsa, M. Li, H. Pobloth, and S. Ragot, "The FlexCode Speech and Audio Coding Approach," *8th ITG Conference Speech Communication*, Aachen, Germany, pp. 1–4, Oct. 2008.

[48] *Audio Codec Processing Functions; Extended Adaptive Multi–Rate – Wideband (AMR–WB+) Codec; Transcoding Functions*, 3GPP TS 26.290, 2004.

[49] R. Lefebvre and C. Laflamme, "Spectral amplitude warping (SAW) for noise spectrum shaping in audio coding," in Proc. *Int. Conf. Acoustic, Speech and Signal Processing (ICASSP'97)*, Munich, vol. 1, pp. 335–338, 1997.

[50] X. Maitre, "7 kHz audio coding within 64 kbit/s," *IEEE Trans. on Selected Areas in Communications*, vol. 6, no. 2, pp. 283–298, Feb. 1988.

[51] B. Kovesi, S. Ragot, and A. Le Guyader, "Coding of digital audio signals," FR200700557792 patent application.

[52] J. Lapierre, R. Lefebvre, B. Bessette, V. Melanovsky, and R. Salami, "Noise shaping in an ITU–T G.711–interoperable embedded codec", Proc. *EUSIPCO*, Lausanne, Switzerland, Aug. 2008.

[53] M. Krasner, M. Berouti, and J. Makhoul, "Stability analysis of APC systems," in Proc. *Int. Conf. Acoustic, Speech and Signal Processing (ICASSP)*, Atlanta, GA, vol. 6, pp. 611–614, 1981.

[54] S. Ragot, B. Kovesi, and A. Le Guyader, "Controlling a noise–shaping feedback loop in a digital audio signal encoder," FR20100055037 patent application.

[55] B. Kövesi, S. Ragot, and A. Le Guyader, "An 64–80–96 kbit/s scalable wideband speech coding candidate for ITU–T G.711–WB standardization," in Proc. *Int. Conf. Acoustic, Speech and Signal Processing (ICASSP'08)*, Las Vegas, NV, pp. 4801–4804, 2008.

[56] S. Ragot et al, "ITU–T G.729.1: An 8–32 kbit/s scalable coder interoperable with G.729 for wideband telephony and voice over IP," *Proc. ICASSP*, Honolulu, vol. 4, pp. 529–532, Apr. 2007.

[57] B. Geiser, S. Ragot, and H. Taddei, "Embedded speech coding: from G.711 to G.729.1," Chapter 8 in *Advances in Digital Speech Transmission*, (R. Martin, U. Heute, C. Antweiler, eds.), New York: Wiley, Jan. 2008, pp. 201–248.

[58] Y. Hiwasaki et al, "G.711.1: a wideband extension to ITU–T G.711," *Proc. EUSIPCO*, Lausanne, Switzerland, Aug. 2008

[59] L. Miao et al, "G.711 Annex D and G.722 Annex B–New ITU–T superwideband codecs," in Proc. *Int. Conf. Acoustic, Speech and Signal Processing (ICASSP)*, Prague, pp.5232–5235, 2011.

[60] B. Kövesi, S. Ragot, and A. Le Guyader, "Encoding of an audio signal with noise transformation in a scalable encoder," FR20080057839 patent application.

[61] B. Kövesi, S. Ragot, and A. Le Guyader, "Encoding with noise shaping in a hierarchical encoder," FR20100053851 patent application.

[62] B. Kövesi et al, "Re–engineering ITU–T G.722: Low delay and complexity superwideband coding at 64 kbit/s with G.722 bitstream watermarking," in Proc. *Int. Conf. Acoustic, Speech and Signal Processing (ICASSP)*, Prague, pp. 5248–5251, 2011.

## B iographies

**Stéphane Ragot** (stephane.ragot@orange.com) received his diplôme d´ingénieur in telecommunications engineering from Telecom Bretagne, France, in 1997. He received his MSc and PhD degrees in electrical engineering from the University of Sherbrooke, Canada, in 2000 and 2003. From 1997 to 2003, he was a research assistant at the University of Sherbrooke. From 2000 to 2003, he was a research engineer at VoiceAge, Canada. Since 2003, he has been with France Telecom R&D/Orange Labs, France. He has contributed to the standardization of speech/audio coders in 3GPP and ITU–T. Since 2008, he has been vice chair of 3GPP SA4. His main research interests include source coding and speech/audio processing.

**Balázs Kövesi** (balazs.kovesi@orange.com) received his degree in electrical engineering from the Technical University of Budapest in 1992. He received his MSc degree from Telecom Bretagne, France, in 1993 and his PhD degree from the University of Rennes I, France, in 1997. He joined the Speech and Audio Coding Group of France Telecom/Orange as a postdoctoral fellow in 1997 and as a research engineer in 1998. His main research interests include speech and audio compression.

**Alain Le Guyader** (alain–le–guyader@orange.fr) received his doctorate in electronic engineering from Rennes University, Rennes, France, in 1978. In 1977, he joined CNET/France Telecom R&D, France. His main research interests include speech and audio coding and audio watermarking. Since 2009, he has been a part–time lecturer in speech and audio coding at the University of Rennes 1/ENSSAT, France.