Parametric stereo extension of ITU-T G.722 based on a new downmixing scheme

Thi Minh Nguyet HOANG [1], Stephane RAGOT [1], Balazs KÖVESI [1], Pascal SCALART [2]

[1] *Orange Labs /TECH/OPERA/TPS, 2 Pierre Marzin, 22307 Lannion Cedex, France*
{*firstname.lastname*}@orange-ftgroup.com

[2] *ENSSAT/IRISA, University of Rennes 1, 6 rue de Kerampont, 22300 Lannion, France*
pascal.scalart@enssat.fr

Abstract—**In this paper, we present a novel, frequency-domain stereo to mono downmixing, which preserves the energy of spectral components and avoids setting the left or right channel as a phase reference. Based on this downmixing technique, a parametric stereo analysis-synthesis model is described in which subband stereo parameters consist of interchannel level differences and phase differences between the mono signal and one of the stereo channels (left or right). This model is applied to the stereo extension of ITU-T G.722 at 56+8 and 64+16 kbit/s with a frame length of 5 ms. AB test results are provided to assess the quality of the proposed downmixing technique. In addition, the quality of the proposed G.722-based stereo coder is compared against reference coders (G.722.1 at 24 and 32 kbit/s dual mono and G.722 at 64 kbit/s dual mono) for clean speech, noisy speech and music.**

I. Introduction

Stereo is widely used in audio applications such as streaming, broadcasting or storage, and significant progress was made in reducing the bit rate for (joint) stereo coding, as shown by the evolution of MPEG audio standards (MP3, AAC, HE-AAC, USAC). On the other hand, in conversational applications speech coders are designed to handle mostly mono signals; stereo, when supported by the service (e.g conferencing), is usually coded using dual mono, that is by coding separately each channel [1]. Recently, ITU-T SG16 has launched several standardization activities aiming at extending existing wideband (50-7000 Hz) mono coding standards to superwideband (50-14000 Hz) and stereo. Examples are given by G.729.1-SWB [2], G.718-SWB [3], and G722/G.711.1-SWB [4]. In these examples, the bitrate set for stereo does not allow dual mono coding and therefore joint stereo coding operating at lower bit rate than dual mono is needed. The present work focuses on the G.722/G.711.1-SWB activity, and presents an experimental stereo extension of G.722 that follows the constraints given in [4] for the stereo extension, e.g. frame length of 5 ms and additional bit rate of 8 or 16 kbit/s.

Let alone dual mono coding, classical techniques for stereo coding are mid/side (M/S) and intensity stereo (IS) coding [5], and more recently parametric techniques such as Binaural Cue Coding (BCC) [6], [7], [8] and Parametric Stereo (PS) coding [9]. The most efficient approach – BCC and PS coding – consists in representing the stereo signal as a mono signal (obtained by stereo to mono downmixing) together with some side information describing the spatial image as it is perceived by the human auditory system. The side information is usually a combination of short-term stereo parameters (or cues) defined per frequency subband [9]:

- Inter-channel Level Difference (ICLD) measuring the level difference (or balance) between channels,
- Inter-channel Time Difference (ICTD) or Inter-channel Phase Difference (ICPD) describing respectively the time or phase difference between channels,
- Inter-channel Coherence (ICC) which represents the coherence (or amount of correlation) between channels.

The above parameters are used at the decoder to control the stereo synthesis that will upmix the mono channel to reconstruct a spatial impression similar to the original one.

Note that for conversational applications some low-delay stereo coders were proposed based on linear prediction techniques [10], [11]. Still, these techniques do not exploit efficiently the above perceptual cues. For this reason, in this work an approach similar to BCC and PS coding was selected to code the perceptually relevant information necessary to extend G.722 in stereo.

This paper is organized as follows. First stereo to mono downmixing is reviewed and a novel downmixing technique in frequency domain is presented in Sec. II. In Sec. III, a parametric stereo analysis-synthesis based on the proposed downmixing scheme is described. An application to the stereo extension of ITU-T G.722 is discussed in Sec. IV. In Sec. V, experimental results are presented before concluding.

II. Stereo to mono downmixing

Many downmixing techniques in time or frequency domain have been developed [12], [13], [7]. Downmixing in time domain does not control finely the phase differences between channels, and make it difficult to preserve the energy per frequency regions. Downmixing in frequency domain can avoid theses disadvantages, however this approach comes with some extra delay and complexity due to the use of time/frequency transforms.

MMSP'10, October 4-6, 2010, Saint-Malo, France.
???-?-????-????-?/10/$??.?? ©2010 IEEE.

A. Review existing stereo to mono downmixing techniques

Two types of downmix can be distinguished: a passive downmix corresponds to a direct matrixing of stereo channels; and an active downmix includes energy and/or phase control.

A general downmixing technique in complex frequency domain (after Fourier analysis) is described in [13] where the mono signal $M[j]$ is obtained by a linear combination of Left (L) and Right (R) channels as follows:

$$M[j] = \omega_1 L[j] + \omega_2 R[j] \tag{1}$$

wherein ω_1, ω_2 are complex values, j corresponds to the index of frequency coefficient. If $\omega_1 = \omega_1 = 0.5$, the mono signal corresponds to an averaging of the two channels.

A special case of this downmix was proposed in [12] where the L and R channels are aligned before downmixing. The L channel for each subband is chosen as the phase reference, the R channel is aligned according to the phase of the L channel by the following formula:

$$R'[j] = \exp^{jICPD[b]} .R[j] \tag{2}$$

where $R'[j]$ is the aligned R channel, j is index of the coefficient in the b^{th} frequency subband. ICPD per frequency subband is defined as follows:

$$ICPD[b] = \angle(\sum_{j=k_b}^{k_{b+1}-1} L[j].R^*[j]) \tag{3}$$

where $[k_b, k_{b+1}]$ are the frequency boundaries of the corresponding subband and $*$ denotes the complex conjugate. The mono signal is then computed by averaging the L and aligned R' channels [12]:

$$M[j] = \frac{L[j] + R'[j]}{2} \tag{4}$$

Using the phase of L channel to align the R channel, signal cancellation can be avoided in the case of out-of-phase channels. Yet, the downmix depends completely on the channel that is chosen as phase reference. For a stereo signal where the L channel (phase reference) has a phase which is not well conditioned (e.g. zero or low-level noise channel), the mono signal does not preserve well the components of the stereo signal.

B. Proposed new downmixing technique

We propose a novel downmixing technique in complex frequency domain to avoid some drawbacks of the downmixing technique of [12].

The stereo channels $L[j]$ and $R[j]$ are downmixed to mono signal $M[j]$ according to the following formula:

$$M[j] = |M[j]| \cdot \exp^{j\angle M[j]} \tag{5}$$

where $|M[j]|$ and $\angle M[j]$ representing respectively magnitude and phase for each frequency subband, defined as:

$$\begin{cases} |M[j]| = \dfrac{|L[j]| + |R[j]|}{2} \\ \angle M[j] = \angle (L[j] + R[j]) \end{cases} \tag{6}$$

This downmixing method is illustrated in Fig. 1. The downmix is computed per frequency bin assuming a complex Fourier analysis of the stereo signals. The magnitude of $M[j]$ is the average of L and R magnitudes. The phase of $M[j]$ is given by the phase of L+R.

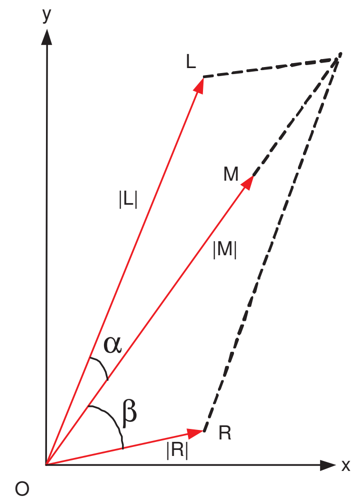


Fig. 1. Proposed downmix in complex frequency domain.

The proposed downmixing technique preserves the energy of the mono signal, as in [12], and avoids the dependency on one arbitrary reference channel to define the phase of the mono signal. By definition, the phase of $M[j]$ lies in the interval delimited by $\angle L[j]$ and $\angle R[j]$. The extreme cases $\angle M[j] \approx \angle L[j]$ or $\angle M[j] \approx \angle R[j]$ are respectively found when L or R is dominant ($|L[j]/R[j]| >> 1$ or $<< 1$).

III. Stereo analysis-synthesis based on the proposed downmixing scheme

A block diagram of parametric stereo analysis-synthesis is shown in Fig. 2. This scheme is based on Fourier analysis-synthesis for time-frequency (T/F) and frequency-time (F/T) transforms. We describe in this section the blocks corresponding to the stereo parameter extraction and stereo synthesis. The block denoted 'DMX' corresponds to the downmixing described in Sec. II-B.

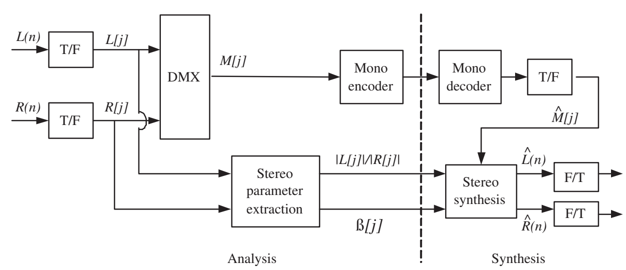


Fig. 2. Parametric stereo analysis-synthesis.

To simplify the presentation of the analysis-synthesis model, the stereo parameter are not quantized – the quantization is considered in Sec. IV where this model is applied to a real coder. Furthermore we assume here that frequency subbands are limited to one frequency bin; the use of frequency subbands of unequal sizes is considered in Sec. IV.

A. Extraction of stereo parameters

Based on the spectra $L[j]$ and $R[j]$, the dominant channel $X[j]$ and the secondary channel $Y[j]$ per frequency bins are determined as:

$$\begin{cases} X[j] = L[j] \\ Y[j] = R[j] \end{cases} \quad if \quad \frac{|L[j]|}{|R[j]|} \geq 1 \tag{7}$$

and

$$\begin{cases} X[j] = R[j] \\ Y[j] = L[j] \end{cases} \quad if \quad \frac{|L[j]|}{|R[j]|} < 1 \tag{8}$$

The complex vectors associated with the dominant channel $X[j]$ and the secondary channel $Y[j]$ are illustrated in Fig. 3 where the phases $\alpha[j]$ and $\beta[j]$ are also defined with respect to the mono signal $M[j]$ as follows:

$$\begin{cases} \beta[j] = \angle(Y[j].M^*[j]) \\ \alpha[j] = \angle(X[j].M^*[j]) \end{cases} \tag{9}$$

To simplify the notations, the index of frequency coefficients do not appear in this figure. Thus $X[j]$, $Y[j]$ and $M[j]$ are denoted X, Y and M, respectively.

At the analysis side, for each frequency bin the extracted stereo parameters are :

- The magnitude ratio $\frac{|L[j]|}{|R[j]|}$ between the two stereo channels, which is equivalent to the ICLD parameter,
- The phase difference ($\beta[j]$) between the secondary channel and the mono signal - see Fig. 2.

Without loss of generality we assume here that $\alpha[j]$ is not computed and only $\beta[j]$ is kept for the synthesis. The missing parameter $\alpha[j]$ will be inferred at the synthesis side with no side information, based on $\beta[j]$ and $\frac{|L[j]|}{|R[j]|}$.

B. Synthesis of stereo parameters

Note that all signals in this section are denoted by $\hat{}$ to differentiate parameters used at the synthesis side. The stereo synthesis reconstructs the L and R channels using the decoded mono signal $\hat{M}(n)$ (transformed in frequency domain) and the parameters extracted at the analysis side.

If only the ratio $\widehat{\frac{|L[j]|}{|R[j]|}}$ per frequency bin is available, the L and R channels are synthesized by adjusting the level of $\hat{M}[j]$:

$$\begin{cases} \hat{L}[j] = c_1[j].\hat{M}[j] \\ \hat{R}[j] = c_2[j].\hat{M}[j] \end{cases} \tag{10}$$

wherein $c_1[j]$ and $c_2[j]$ are defined as follows:

$$\begin{cases} c_1[j] = \frac{2\hat{I}[j]}{\hat{I}[j]+1} \\ c_2[j] = \frac{2}{\hat{I}[j]+1} \end{cases} \tag{11}$$

with $\hat{I}[j] = \widehat{\frac{|L[j]|}{|R[j]|}}$. The above synthesis is equivalent to intensity stereo coding, where only ICLD stereo parameters are available.

If both $\widehat{\frac{|L[j]|}{|R[j]|}}$ and $\hat{\beta}[j]$ are available, the stereo synthesis is given by:

$$\begin{cases} \hat{X}[j] = max(c_1[j], c_2[j]).\hat{M}[j].\exp^{j\hat{\alpha}[j]} \\ \hat{Y}[j] = min(c_1[j], c_2[j]).\hat{M}[j].\exp^{j\hat{\beta}[j]} \end{cases} \tag{12}$$

The stereo channels $\hat{R}[j]$ and $\hat{L}[j]$ are derived from $\hat{X}[j]$ and $\hat{Y}[j]$ using $\hat{I}[j]$ as in Eq. 7. The condition on $\hat{I}[j]$ is necessary to distinguish the dominant ($\hat{X}$) and secondary ($\hat{Y}$) channels, and match the angles $\hat{\alpha}[j]$ and $\hat{\beta}[j]$ with the respective channels.

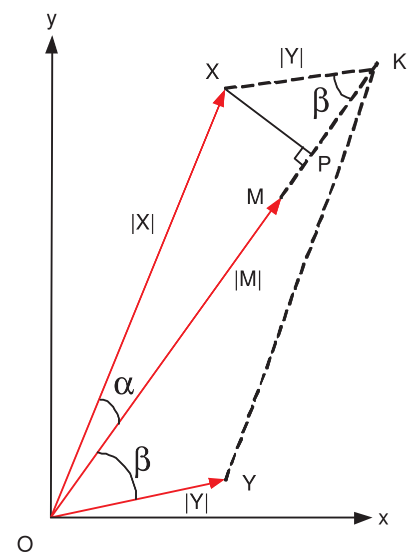


Fig. 3. Geometry of the proposed downmix and related properties.

In the following we show how the angle $\hat{\alpha}[j]$ can be estimated based on the available parameters: $\widehat{\frac{|L[j]|}{|R[j]|}}$ and $\hat{\beta}[j]$. The estimation of $\hat{\alpha}[j]$ is illustrated in Fig. 3.

If we project the dominant channel $X[j]$ onto the line OM, where O and M correspond to the zero complex value and $M[j]$ respectively, we obtain P with:

$$|XP| = \left|\hat{X}[j]\right|.|\sin\hat{\alpha}[j]| = \left|\hat{Y}[j]\right|.\left|\sin\hat{\beta}[j]\right| \tag{13}$$

The phase $\hat{\alpha}[j]$ is therefore obtained as follows:

$$\hat{\alpha}[j] = \pm \arcsin\left(\frac{\left|\hat{Y}[j]\right|}{\left|\hat{X}[j]\right|}.\left|\sin\hat{\beta}[j]\right|\right) \tag{14}$$

The sign of $\hat{\alpha}[j]$ is determined by observing that $\hat{\alpha}[j]$ and $\hat{\beta}[j]$ are opposite.

IV. Application to stereo extension of ITU-T G.722

ITU-T SG16 has recently launched a standardization activity, G.722-SWB (within the Q.10/16), which consists in extending the G.722 recommendation in two ways:

1) An extension of the acoustic band from wideband (50-7000 Hz) to superwideband (50-14000 Hz).
2) An extension from mono to stereo.

Two wideband stereo extension modes of G.722 envisioned:

- A stereo extension of G.722 at 56 kbit/s with an additional bitrate of 8 kbit/s, giving 64 kbit/s in total.
- A stereo extension of G.722 at 64 kbit/s with an additional bitrate of 16 kbit/s, giving 80 kbit/s in total.

We review in this section the G.722 standard and describe a stereo extension of G.722 that is motivated by the G.722-SWB project.

A. Review of ITU-T G.722

The ITU-T G.722 Recommendation is based on a subband embedded ADPCM coding scheme [14], [15]. The input signal of G.722, sampled at 16 kHz, is decomposed into two subbands [0-4000 Hz] and [4000-8000 Hz] by a Quadrature Mirror Filters (QMF). Each subband is coded separately: the lower subband is coded by an embedded ADPCM coder at 6, 5 and 4 bits per sample while the higher subband is coded by an ADPCM coder with 2 bits per sample. The total bit rate is 64, 56 or 48 kbit/s according to the number of bits allocated to the lower subband. Fig. 4 presents the block diagram of G.722.

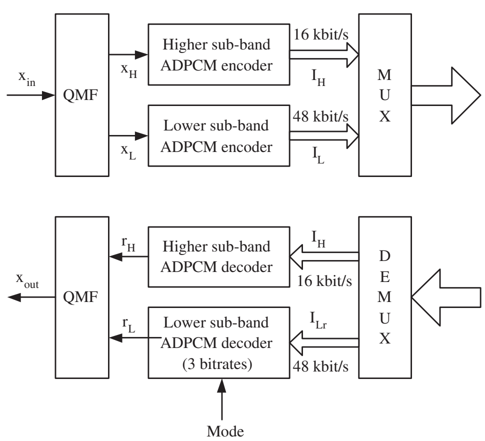


Fig. 4. Block diagram of the 48, 56 and 64 kbit/s (7 kHz) G.722 coder.

A bitstream of G.722 can be viewed as a sequence of (scalar) quantization indices coded with 8, 7 or 6 bits per sample.

G.722 has an algorithmic delay of 22 samples due to the QMF filterbank and can operate with any frame size of $2n$ samples, $n \geq 1$. In this work G.722 is used with a frame size of 5 ms (80 samples).

B. Description of the proposed stereo encoder

A block diagram of the proposed encoder is shown in Fig. 5. The Left (L) and Right (R) channels are deinterleaved and processed by a high-pass filter (HPF) to remove the components below 50 Hz. Each channel is then transformed in frequency domain by short-term Fourier analysis. A sine window of 160 samples (10 ms) is used with 50% overlap; in other words, the signal is weighted by an analysis window that covers including the current frame (5 ms) and the next frame (lookahead of 5 ms). The resulting spectra, $L'[j]$ and $R'[j]$ $(j = 0 \ldots 80)$, comprise 81 complex coefficients, with a resolution of 100 Hz per frequency bin.

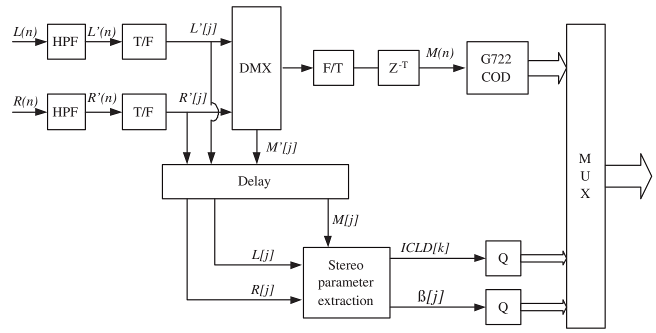


Fig. 5. Proposed G.722 stereo encoder.

The spectra $L'[j]$ and $R'[j]$ are partitioned into 20 frequency subbands following approximately a Bark scale. Note that the first subbands comprise only one frequency bin. The extraction of stereo parameters is the same as in Sec. III-A, except that the ICLD (ratio $L[j]/R[j]$) is computed per subband – this extraction is identical for the first subbands reduced to one frequency bin.

With 5 ms frames the stereo extension has a limited bit budget per frame: 40 bits at 8 kbit/s and 80 bits at 16 kbit/s. The quantization of stereo parameters is therefore optimized for the constraint of short frame size and low bit budget.

In the first layer of stereo extension (+8 kbit/s), the ICLD parameters are coded by a differential non-uniform scalar quantization using 40 bits per frame. To make it possible, the ICLD parameters are partitioned into two blocks of 10 subbands and quantized by alternative the coded blocks in each frame; hence in frames of even index t only the first block is coded, while in frames of odd index t only the second block is coded. This is equivalent to refreshing the coded block every two frames (10 ms).

A second layer of stereo extension (+8 kbit/s) coming on top of the first layer (i.e. hierarchical) represents the phase information $\beta[j]$ with 40 bits; note that the phase is coded only in the most important frequency bins (< 1 kHz). As described in Sec. III-A, only the phase difference $\beta[j]$ between the dominant channel $X[j]$ and the mono signal $M[j]$ is coded. The phase $\beta[j]$ is quantized with 5 bits within the interval $[-\pi, \pi]$ using a uniform quantizer of stepsize $\pi/16$ bits for $j = 2, \ldots, 9$.

C. Bit allocation of the proposed coder

Tab. I and Tab. II present the bit allocation for the stereo extension at 56+8 and 64+16 kbit/s respectively and the bits distribution for each stereo parameter, where k is the subband index. Note that in even frames $ICLD[k = 0]$ is not coded.

D. Description of the proposed stereo decoder

A block diagram of the proposed stereo decoder is shown in Fig. 6. The part of bitstream corresponding to G.722 (Layer 0 is decoded at 56 or 64 kbit/s according to the G.722 mode.

TABLE I
BIT ALLOCATION FOR THE LAYER 56+8 KBIT/S.

Frame	Parameters	Bit allocation
even	ICLD[k=1,..,9]	5 + 8x4 = 37 bits (3 unused bits)
odd	ICLD[k=10,..,19]	5 + 8x4 + 3 = 40 bits

TABLE II
BIT ALLOCATION FOR THE LAYER 64+16 KBIT/S.

Frame	Parameters	Bit allocation
even	ICLD[k=1,..,9], $\beta[j]$	37 bits for ICLD 8x5 = 40 bits for phase
odd	ICLD[k=10,..,19], $\beta[j]$	40 bits for ICLD 8x5 = 40 bits for phase

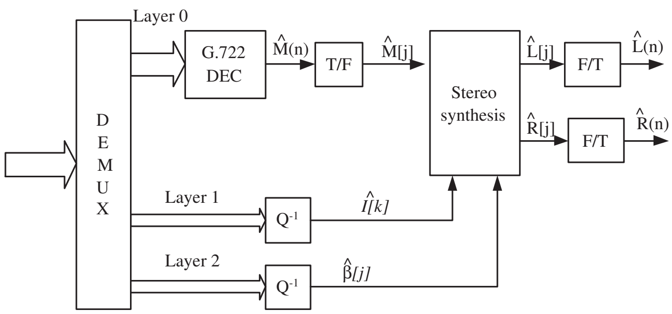


Fig. 6. Proposed G.722 stereo decoder.

At 56+8 kbit/s the ICLD parameters (per subband) are decoded and the synthesis is performed according to Eq.10.

At 64+16 kbit/s the ICLD parameters (per subband) and the phase difference $\beta[j]$ (per frequency bin) are decoded. The L and R channels are synthesized in the same manner as Eq. 12.

V. EXPERIMENTAL RESULTS

Two subjective listening tests were conducted:

1) An AB test to compare the quality of the proposed downmixing with the method of [12].
2) A MUSHRA test [16] to evaluate the quality of the proposed stereo coder compared to reference coders.

In total, 12 test items were used as listed in Tab. III – they were all P.341 pre-filtered to the 50-7000 Hz bandwidth and normalized in level. There were 7 expert listeners for AB test and 5 expert listeners for MUSHRA test. The tests were conducted using high-quality headphones. The speech signals were in French language with single (male/female) talkers and double talkers.

TABLE III
TABLE OF TEST ITEMS.

Category	Item	Description
clean speech	t1s1, t3s3 t2s1, t4s2	binaural reverberant MS anechoic
noisy speech	t7s2, t8s1 t5s2, t6s1	interf. talker with bin. mic office noise with MS mic
music	music5_inv, music6 music8, music9	the Beatles, choral song rap, classical latin

A. AB test results (downmixing)

The AB test is conducted using the grading scale shown in Tab. IV.

TABLE IV
TABLE OF COMPARISON FOR AB TEST.

-3	A significantly worse than B
-2	A worse than B
-1	A slightly worse than B
0	A equivalent to B
1	A slightly better than B
2	A better than B
3	A significantly better than B

The test results in Fig. 7 show that the quality of the proposed downmixing technique is slightly better on average than the downmixing mono proposed in [12]. For one specific item (music5_inv) the quality difference is very significant. For this item the downmixing technique in [12] does not perform well because the phase of the L channel is partly random.

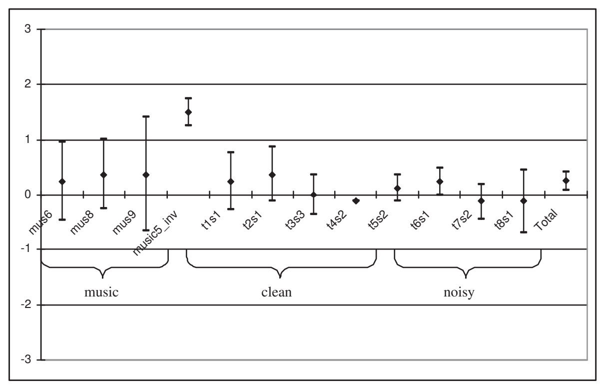
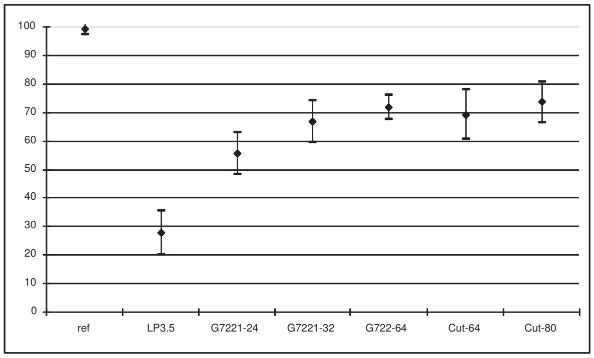


Fig. 7. AB test results of downmixing methods (A = proposed downmixing technique, B = reference [12]).

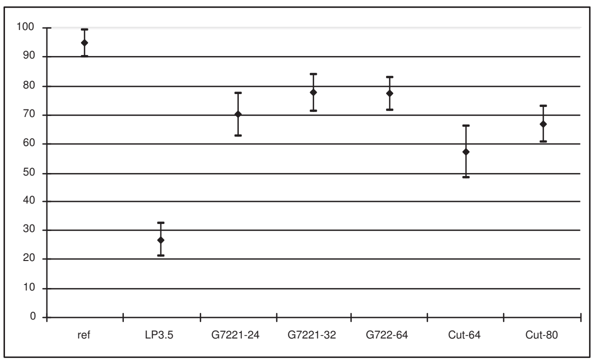
B. Quality results for stereo coders

The coders used in the MUSHRA test are: G.722.1 dual mono (2x24 and 2x32 kbit/s) with frames of 20 ms, G.722 dual mono (2x64 kbit/s) with frames of 5 ms, and the proposed stereo coder denoted CuT (Coder under Test) with frames of 5 ms. Note that the G.722.1 coder [17] is a wideband (50-7000 Hz) transform coder based on the Modulated Lapped Transform (MLT), operating with frames of 20 ms overlapping by 50%; its performance is known to be good for music, reverberant speech and most types of noisy speech, but it suffers from pre-echo artefacts for clean speech.

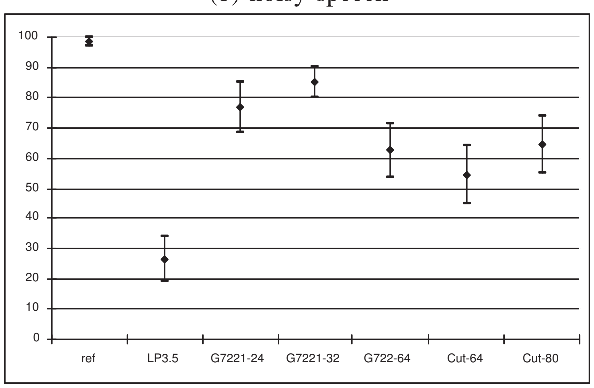
The MUSHRA test results are shown in Fig. 8. The relative performance of the proposed stereo coder at 56+8 and 64+16 kbit/s can be easily explained. At 56+8 kbit/s, only ICLD parameter is used to synthesize the stereo channels, therefore the stereo image is limited for stereo signals that were not produced by simple pan-pot; in such cases the stereo synthesis fails to capture the spatial width or phase/time differences. At 80 kbit/s, the L and R channels are synthesized with the ICLD and phase parameters, the stereo channels get closer to the original stereo signal.



(a) clean speech



(b) noisy speech



(c) music

Fig. 8. MUSHRA test results for each category (clean speech, noisy speech, music).

The results for clean speech show that the quality of the proposed coder at 56+8 and 64+16 kbit/s is equivalent to G.722 at 64 kbit/s dual mono. The advantage of the proposed coder is that it operates at half bitrate compared to dual mono (2 x G.722 at 64 kbit/s). The performance difference with G.722.1 dual mono is mostly due to the pre-echo artefacts with G.722.1.

For noisy speech, the quality of the proposed coder is lower than G.722.1 and G.722 dual mono, except at 64+8 kbit/s where the quality is equivalent to G.722.1 at 24 kbit/s dual mono. The difference with G.722 dual mono can be attributed to the limited spatial width and fidelity.

For music, the quality of the proposed stereo coder is equivalent to the G.722 dual mono but inferior to the G.722.1 dual mono. G.722.1 performs well for music while there are some limitations in the proposed stereo coder (G.722 introducing noise in 4000-8000 Hz, limited spatial width).

VI. Conclusion

In this paper, we proposed a stereo to mono downmixing technique in frequency domain. This technique preserves the energy of mono signal and avoids issues due to the complete dependency on one channel (L or R) for the phase computation. A parametric stereo extension of G.722 at 56+8 and 64+16 kbit/s was described; the quality of the proposed coder was evaluated in a MUSHRA test. The proposed stereo coder operates at the lower bitrate than G.722 dual mono, with a speech and music quality at 64+16 kbit/s that is equivalent to G.722 dual mono. Future work will aim at improving quality, especially for noisy speech and music.

References

[1] A. Nagle, C. Quinquis, A. Sollaud, A. Battistello, and D. Slock, "Quality Impact of Diotic versus Monaural Hearing on Processed Speech," in *Preprint 7220, 123th AES convention*, New York, Oct. 2007.

[2] ITU-T TD109R1 (WP3/16) Annex Q10.C, "Terms of Reference (ToR) for the super wideband extension to ITU-T G.729.1," study Period 2009–2012, Geneva, Nov. 2009 (Source: Q.10/16 Rapporteurs).

[3] ITU-T TD109R1 (WP3/16) Annex Q10.E, "G.718 SWB ToR," study Period 2009–2012, Geneva, Nov. 2009 (Source: Q.10/16 Rapporteurs).

[4] ITU-T TD109R1 (WP3/16) Annex Q10.D, "Terms of Reference (ToR) and time schedule for the super wideband extension to ITU-T G.722 and ITU-T G.711.1," study Period 2009–2012, Geneva, Nov. 2009 (Source: Q.10/16 Rapporteurs).

[5] J. Herre, "From Joint Stereo to Spatial Audio Coding - Recent Progress and Standardization," in *6th International Conference on Digital Audio Effects (DAFX04)*, Naples, Italy, Oct. 2004.

[6] C. Faller and F. Baumgarte, "Efficient Representation of Spatial Audio Using Perceptual Parametrization," in *Proc. IEEE Workshop on Appl. of Sig. Proc. to Audio and Acoust.*, New York, Oct. 2001.

[7] C. Faller, "Parametric Coding of Spatial Audio," Ph.D. dissertation, Ecole Polytechnique Federale de Lausanne, 2004.

[8] C. Faller and F. Baumgarte, "Binaural Cue Coding - Part ii: Schemes and Applications," *IEEE Trans. on Speech and Audio Proc.*, vol. 11, no. 6, pp. 520–531, Nov. 2003.

[9] E. Schuijers, W. Oomen, B. Brinker, and J. Breebaart, "Advances in Parametric Coding for High-Quality Audio," in *Preprint 5852, 114th AES convention*, Amsterdam, Mar. 2003.

[10] H. Kruger and P. Vary, "A New Approach for Low-Delay Joint-Stereo Coding," in *ITG-Fachtagung Sprachkommunikation*, Aachen, Oct. 2008.

[11] S. A. Ramprashad, "Stereophonic CELP coding using cross channel prediction," in *Proceedings. 2000 IEEE Workshop on Speech Coding*, Delavan, Sep. 2000.

[12] Samsudin, E. Kurniawati, N. B. Poh, F. Sattar, and F. George, "A Stereo to Mono Dowmixing Scheme for MPEG-4 Parametric Stereo Encoder," in *Proc. ICASSP'06*, Toulouse, May 2006.

[13] J. Breebaart, S. van de Par, A. Kohlrausch, and E. Schuijers, "Parametric coding of stereo audio," *EURASIP Journal on Applied Signal Processing*, vol. 2005, pp. 1305–1322, Jan. 2005.

[14] X. Maitre, "7 kHz Audio Coding within 64 kbit/s," *IEEE Journal on Selected Areas in Communications*, vol. 6, pp. 283–298, Feb. 1988.

[15] P. Mermelstein, "G.722: a new CCITT coding standard for digital transmission of wideband audio signals," *IEEE Communications Magazine*, vol. 26, pp. 8–15, Jan. 1988.

[16] ITU-R Rec. BS.1534, "Method for the subjective assessment of intermediate quality level of coding systems," Mar. 2003.

[17] ITU-T Rec. G.722.1, "Coding at 24 and 32 kbit/s for hands-free operation in system with low frame loss," Sept. 1999.