Single-Mode-Based Unified Speech and Audio Coding by Extending the Linear Prediction Domain Coding Mode

Seungkwon Beack, Jongmo Seong, Misuk Lee, and Taejin Lee

Unified speech and audio coding (USAC) is one of the latest coding technologies. It is based on a switchable coding structure, and has demonstrated the highest levels of performance for both speech and music contents. In this paper, we propose an extended version of USAC with a single-mode of operation—which does not require a switching system—by extending the linear prediction-coding mode. The main concept of this extension is the adoption of the advantages of frequency-domain coding schemes, such as windowing and transition control. Subjective test results indicate that the proposed scheme covers speech, music, and mixed streams with adequate levels of performance. The obtained quality levels are comparable with those of USAC.

Keywords: USAC, HE-AACv2, AMR-WB+.

I. Introduction

Motivated by different factors and backgrounds, speech and audio coding technologies underwent separate development paths. The main initial motivation of a speech-coding scheme is the successful delivery of information even under low bitrates; therefore, a model-based approach is generally used [1]. However, owing to the limitations of model-based approaches, speech-coding schemes have limited performances when extended to higher bitrates [2]. In contrast, audio-coding schemes have succeeded in achieving higher levels of audial perceptual quality and information delivery. Because such schemes have been developed for human hearing, most of them include a perceptual quantization process within the frequency domain [3]. However, despite their higher audial perceptual characteristics, they cannot appropriately deliver speech information at low bitrates [2], [3].

Unified coding has recently become an interesting solution to both speech and audio compression. One of the remarkable achievements of the unified approach is adaptive multi-rate wide-band plus (AMR-WB+), which adopts a transform-based quantization process for improving audio quality, but suffers from a constraint of interoperability with a speech-coding mode [4], [5]. The latest and most remarkable contributor to speech and audio quality is the unified speech and audio coding (USAC) approach [6], [7]. Instead of enforcing unification, USAC adopts a switching structure and harmonizes existing speech and audio coding technologies. As a result of these harmonization efforts, the coding component of AMR-WB+ has evolved toward a linear prediction domain (LPD) mode, and the advanced audio coding (AAC) components have been modified into a frequency domain (FD) mode. Both LPD and FD modes have been integrated into a unified transform domain—a modified discrete cosine transform (MDCT). USAC has shown superior performance

Manuscript received June 16, 2016; revised Dec. 13, 2016; accepted Jan. 4, 2017. This work was supported by the Institute for Information & Communications Technology Promotion (IITP) grant funded by the Korean government (MSIP) (No. B0101-16-0295; Development of UHD Realistic Broadcasting, Digital Cinema, and Digital Signage Convergence Service Technology).

Seungkwon Beack (corresponding author, skbeack@etri.re.kr), Jongmo Seong (jmseong@etri.re.kr), Misuk Lee (lms@etri.re.kr), and Taejin Lee (tjlee@etri.re.kr) are with the Broadcasting & Media Research Laboratory, ETRI, Daejeon, Rep. of Korea.

when compared with existing state-of-the-art technologies such as AMR-WB+ and high-efficiency AAC version 2 (HE-AACv2) at all evaluated bitrates (from 8-kb/s mono to 96-kb/s stereo) [2], [6]–[8]. To support a switching system, signal classification should be conducted prior to actually encoding the input signal. Signal classification algorithms have been studied [9], [10] and have shown remarkable improvement, but they do not guarantee a desired level of performance when combined with actual audio coding systems. Although the sound quality of both speech and audio systems has dramatically improved, one of the most important remaining issues is how to control the switching system, an encoding issue that is not included within the scope of the moving picture experts group (MPEG) standardization process. This constraint leads to a USAC performance limitation when its switching control is supported by only an ideal signal classifier using an appropriate coding gain method.

In our previous studies, we proposed the use of adaptive windowing with the LPD mode, a technique that showed to be capable of providing a remarkable performance improvement; however, our previous paper did not provide detailed information on its implementation, and was limited to the use of mono [11]. In the current paper, we provide updated and detailed information on the adoption of adaptive-window-based LPD for stereo use at higher bitrates, by resolving related issues such as transient control, mode decision, and stereo extension. The performance is confirmed through listening test results and by comparing them with those of a USAC reference-quality bitstream.

II. USAC Overview

In this section, the overall structure of USAC is discussed [8], [12]. Figure 1 shows a block diagram of the USAC core encoding part; the bandwidth extension and stereo-coding module are not presented, so as to clearly understand the core encoding process, which is considered a part of the current study. USAC is basically a switching system that depends upon the characteristics of the input signals. If the input signals are music-like, the FD coding mode is selected. Before encoding the current input frame in FD mode, the transition property of the following frame should be checked, because the current window type is determined considering the next frame type. The scaling and quantization process in FD mode originally comes from that of AAC. One of the differences is that arithmetic entropy coding is adopted, instead of Huffman coding. On the other hand, when the input signal is speech-like, LPD coding mode is selected. The LPD mode concept derives from the transform-coded excitation coding mode of AMR-WB+, but many components are updated, not only to improve

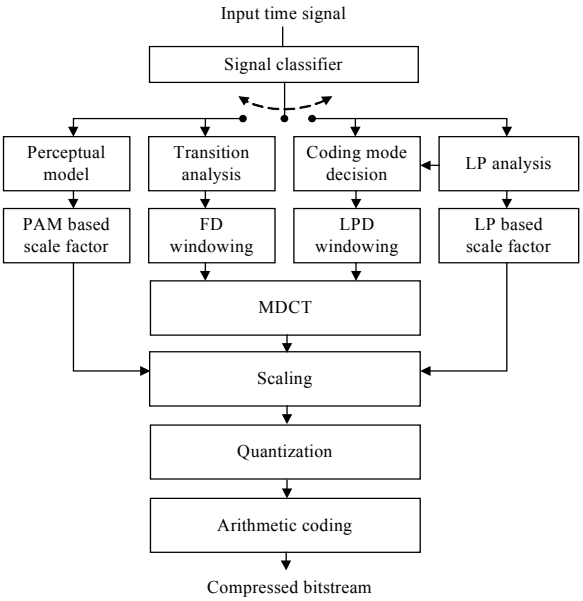


Fig. 1. Encoder structure of USAC.

performance, but also to harmonize the FD mode coding structure. The main updated part is the residual encoding process. A residual signal is commonly obtained in the time domain through a filtering operation using linear prediction coefficients (LPCs); however, in LPD mode, the residual signals are calculated in the MDCT domain by scaling the MDCT coefficients with LPC-based scale factors, which are calculated by converting the time-domain LPCs to their corresponding FD versions, as follows:

$$a_f[k] = \sum_{n=0}^{r-1} a_t[n] \cdot e^{-j\frac{2\pi(k+n/2)}{2M}}, \quad 0 \le k \le M-1, \tag{1}$$

where $a_t[n]$ are the LPCs in the time domain and $a_f[k]$ are the FD coefficients when a discrete Fourier transform is applied. The frequency index k is limited to the number of filter banks M (which is normally set to 64), and index r is the LPC order. A scale factor $g[k]$ is calculated with $1/sqrt\left(a_f[k]a_f^*[k]\right)$ and is multiplied per sub-band. This process is called FD noise shaping, and it allows the USAC structure to obtain a more unified form in the MDCT domain. As a result, the LPD process becomes very similar to the FD process, so as to minimize quantization noise discontinuity when the core encoding mode switches from LPD to FD, or vice versa.

III. USAC Limitations

Although USAC has shown a remarkable level of performance by overcoming the limitations of existing state-of-the-art coding technologies, the confirmed performance is based on the switching system being controlled using signal

classification when operating below 64-kb/s [6]. In this section, we describe the problems inherent to the switching system, as well as the limitations of the current LPD mode.

1. Problems of Switching-Based Coding Systems

USAC is a switching system in which many different technologies have been introduced, aiming to compensate for any distortion or discontinuity when the coding mode is changed. For example, the unification of window-sequence technologies in USAC has been designed to compensate the transition distortion that occurs during the switching process. Forward aliasing cancelation (FAC) technology, which is another compensation approach, was introduced in USAC for transition compensation between ACELP and the MDCT-based coding mode [4]. All transition compensation methods perform well; therefore, USAC achieves thoroughly better performance levels. However, the coding efficiency and sound quality are critically dependent upon the decision of the signal classifier. For instance, if the classifier incorrectly assumes an input speech-like signal to be a music-like signal at a low bitrate, the speech-like input signal may be encoded in FD mode; therefore, better speech quality levels than those of AMR-WB+ cannot be guaranteed. On the other hand, if a music-like signal is encoded using LPD mode, the decoded music signal will not be superior to that of HE-AACv2. Another problem inherent to a switching coding system is that switching may occur abruptly in the middle of the input signals, which may result in ambiguity regarding whether the signals are speech or some other type of signal. This type of situation occasionally occurs for mixed-type input signals such as speech mixed with music, or music mixed with speech. Such abrupt transitions occurring in the middle of the input signals produce spectral discontinuity, despite the use of compensation tools such as FAC or windowing techniques. The internal sampling rate of each core encoding is usually not identical to that of the others. For instance, the core band of FD mode is sampled at 8.0 kHz and 12 kb/s, whereas that of LPD mode is 12.8 kHz. This difference results in an abrupt change in the spectral shape when the coding mode is switched. Figure 2 shows a change in coding-mode in the middle of the signals. The signals in Fig. 2 show strong harmonic components within the low band, and certain parts of the signals are therefore incorrectly output by the classifier as speech; on the other hand, other parts are correctly determined to be music. This means that the signal classifier on the encoder side has to deal with ambiguity (owing to insufficient information) when deciding on the correct classification. As a result, an inappropriate decision boundary creates a spectrogram discontinuity—which can be observed at the switching location—and may degrade sound quality.

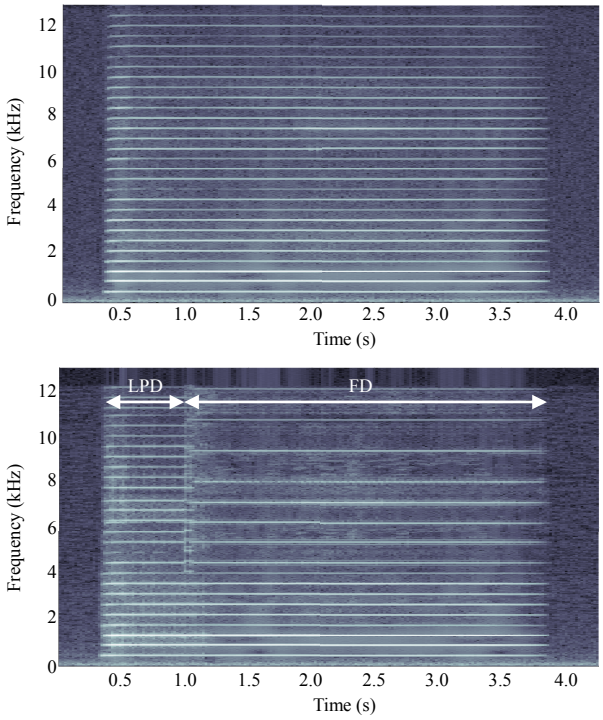


Fig. 2. Spectrogram of an original (upper) and decoded (lower) pitch-pipe waveform using USAC.

2. Windowing Restrictions in LPD Mode

The LPD mode was further updated for improved performance using the coding scheme of AMR-WB+. The main revision was the adoption of MDCT instead of the discrete Fourier transform. The LPD coding mode consists of four parts: LPD80, LPD40, LPD20, and ACELP. With the exception of ACELP, all LPD coding modes are applied in the MDCT domain. The number of coefficients to be quantized in each LPD mode are 1,024, 512, and 256, which correspond to LPD80, LPD40, and LPD20, respectively. All encoding processes in the MDCT domain of the LPD mode are very similar to those of the FD mode. The major windowing restriction in LPD is that the overlap of each LPD coding mode is fixed at 256 samples, as shown in Fig. 3. As a result, the shapes of LPD80 and LPD40 have a flat part in the middle of the window, which can slow down the sidelobe attenuation considerably, and cause inaccurate quantization. Another negative effect of a fixed small overlapping window size is that the discontinuity of the quantization noise between adjacent frames may be audible within the stationary interval of the signal. Therefore, if the windowing process in LPD is revised similarly to that of FD, LPD can mimic FD encoding mode and deal with music-like signals. To revise the windowing

ETRI Journal, Volume 39, Number 3, June 2017
https://doi.org/10.4218/etrij.17.0116.0397

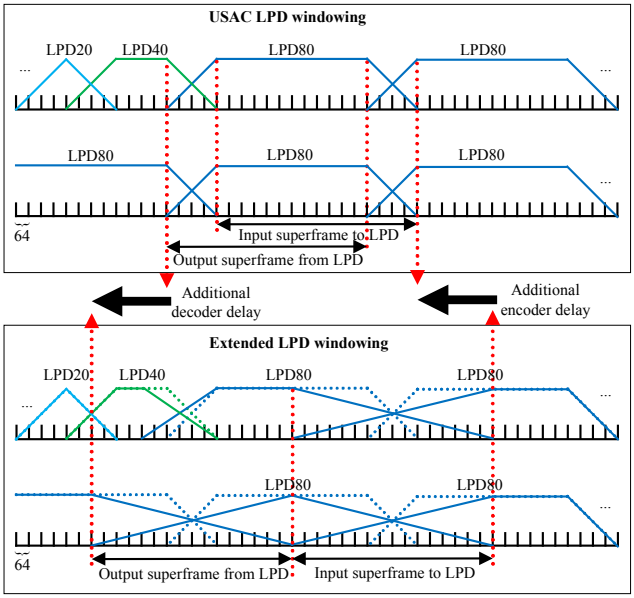


Fig. 3. USAC LPD windowing. Upper figure shows the current USAC LPD windowing, and lower figure shows the extended long-window based LPD windowing.

process, many issues regarding the actual implementation of an LPD extension have been considered in the USAC encoding process. The next section presents detailed information on implementing an extension to the LPD mode.

IV. USAC LPD Mode Extension

Our goal is to obtain single-mode-based USAC operation by removing the dependence on signal classification. The approach that was selected to meet this goal is the development of an LPD extension, because the LPD mode between MDCT-based coding mode (LPD80, LPD40, and LPD20) and ACELP has already been internally unified through the use of FAC technology. Although the FD mode shows a high level of performance for music-like signals, it intrinsically has a performance limitation for speech-like signals [13]. If the LPD mode can adopt the specific advantages of the FD mode, LPD can be expected to show a comparable performance to that of FD for music-like input signals, while preserving the performance obtained with speech-like signals.

1. Adaptive LPD Windowing Extension

The main strength of the FD mode when compared with the LPD mode comes from the shape of the analysis and synthesis window selected for analyzing the spectral components with enhanced frequency selectivity, which can successfully cancel the effect of windowing. We adopted FD windowing and revised the windowing of LPD accordingly [11]. However, to improve their performance, the new window shapes require a further revision of the transition control and buffer delay scheduling aspects. Figure 3 shows the new window sequence of LPD, when the long-overlap-based windowing typical of FD is adopted. The overlap size is maximally extended to 1,024 for the connection of LPD80 to LPD80, and the window shape is perfectly symmetrical, resulting in enhanced frequency selectivity and improved quantization resolution. One of the negative effects of extending the overlap size is the introduction of an additional coding delay of 384 samples at both the encoder and decoder sides; however, this delay is still below the 1,600-sample delay of the FD windowing process, and is required to determine the coding mode of the next superframe. LPD mode is basically determined through the closed-loop-based signal-to-noise ratio [4], [5]. If we adopt long windowing in LPD, the current superframe is encoded by following the mode decision result, which is determined during the previous encoding stage.

2. Transition Strength Determination

To avoid a loss in time resolution because of the windowing process, another consideration should be included in our extension. Common sense implies that long windowing requires large analysis frames, thereby lessening the time resolution required to capture transient behavior of the input signals. The overlap size should therefore be adjusted depending on the variation dynamics of the time signals. The best method for enhancing the time resolution is to replace long windows with shorter ones. Although this process may be considered during the LPD mode decision process (for example, if all LPD coding modes are determined to be LPD20), the replacement of a short window sequence requires more quantization bits to prevent a degradation of the sound fidelity within a stationary interval. For instance, when a transition occurs between adjacent LPD80 modes, the coding mode is intentionally replaced with LPD20 mode for transition encoding; however, the coding gain within the intra-frame may be weakened, because the LPD encoding mode of USAC is selected according to the coding gain within the intra-superframe—by measuring the signal-to-noise ratio of the segment [4]. If we adopt the concept of long overlap windowing,

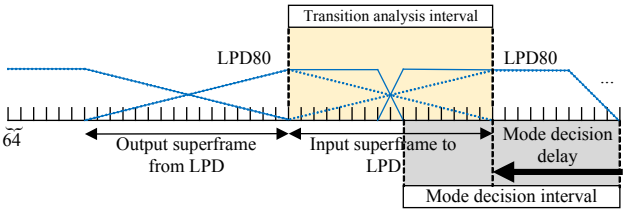


Fig. 4. Transition analysis interval with the LPD extension.

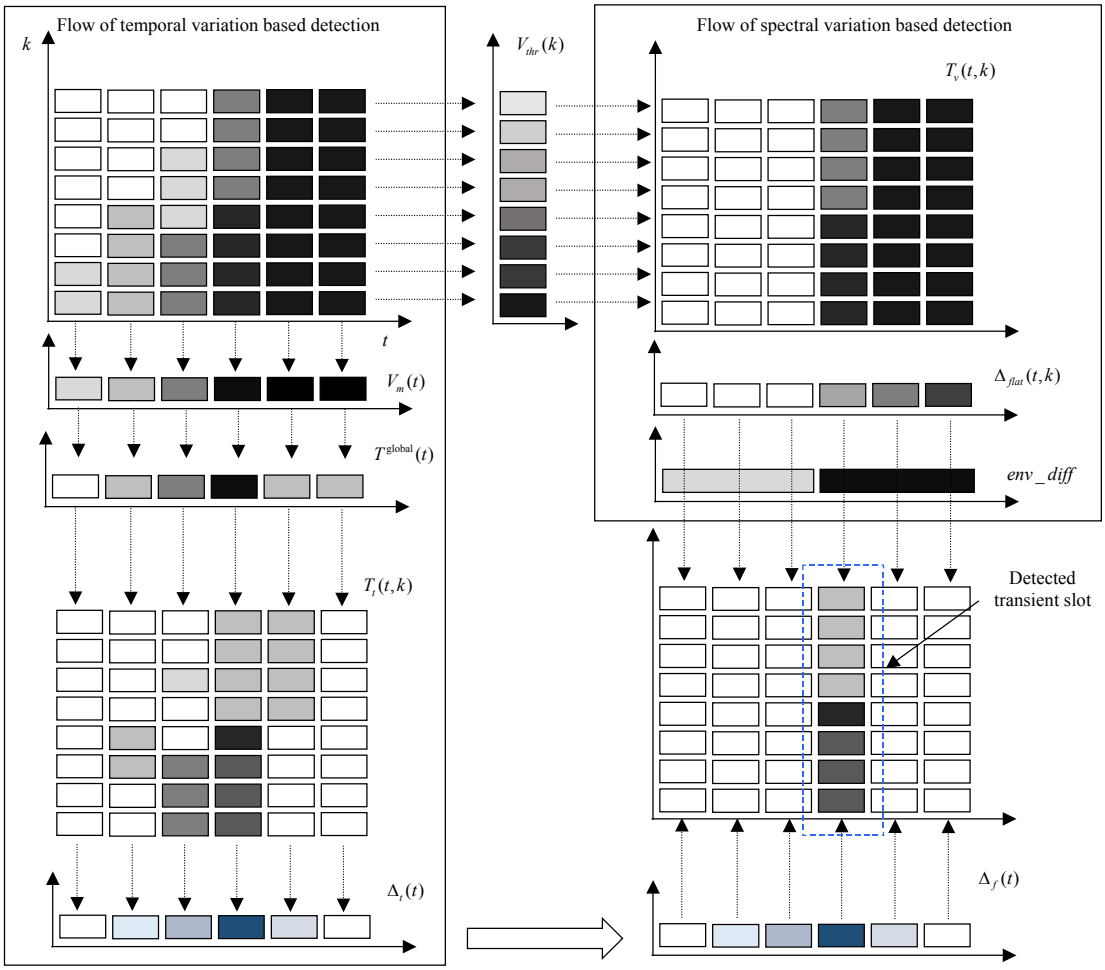


Fig. 5. Transition detection algorithm flow for adaptive windowing in LPD mode.

another approach for achieving a transition between the inter-superframes should be considered. In the proposed LPD mode extension, an inter-frame based transient-analysis mechanism is additionally applied. Figure 4 shows the analysis interval where the strength of the inter-superframes is investigated. When a transition among inter-superframes is observed, the overlap size among the superframes is reduced from 1,024 to 256, or from 512 to 128, depending on the transition strength. The transition-detection algorithm operates in the quadrature mirror filter bank (QMF) domain, which can simultaneously provide temporal and spectral information, thereby allowing the temporal transition to be detected from the changes in the spectral components. A QMF is commonly included in the encoding and decoding processes when a spectral band replication tool is activated, which means that the computation of the QMF conversion for analyzing the interval is not considered. The transition detection algorithm is schematically depicted in Fig. 5.

The blocks in Fig. 5 represent time slot units of the QMF domain. In addition, when 64 QMF bands are applied to the 2,048 input signals, the size of the time slot unit ranges from 1 to 32. Here, k is the index of the sub-bands, t is the index of the time slots, and $V_m(t)$ is the average absolute value of each time block of the input signals, where subscript index m indicates the mid-signal of the stereo inputs. In addition, $T^{global}(t)$ is the threshold sequence calculated by taking the differential value of $V(k) - V(k-1)$, but is set to zero when the differential value is negative. The sequence $T^{global}(t)$ is applied in each sub-band of the corresponding time block, and a $T_t(t, k)$ grid is thus obtained. In addition, $\Delta_t(t)$ is the number of blocks counted in the vertical direction when $T_t(t, k)$ is a positive value. For the other transition detection flow, $V_m(t)$ is calculated as the average absolute value of each sub-band block, and is used as the threshold value to set the sub-band block to zero and obtain the $T_v(t, k)$ grid. After estimating the flatness $\Delta_f(t)$ by considering the vertical statistical variation of $T_v(t, k)$, the position of the envelope variation can be estimated through the difference in $\Delta_{flat}(t)$. The final transition position is determined by monitoring $\Delta_t(t)$ and $\Delta_f(t)$. Normally, each position of $\Delta_t(t)$ and $\Delta_f(t)$ is closely detected, and LPD is switched to short

https://doi.org/10.4218/etrij.17.0116.0397

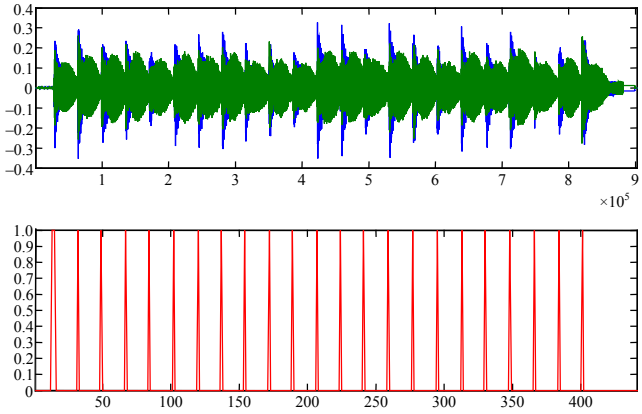


Fig. 6. Example of transient detection; upper figure shows the original stereo signals, and lower figure shows the detected transition position.

windowing. Figure 6 shows the results of the transient position detection when applying the detection algorithm.

3. LPD Stereo Coding Mode Extension

The current USAC reference-quality bitstream does not support LPD-based stereo-coding mode at higher bitrates. The main reason for this is not a restriction of the stereo-coding tool, but rather the fact that the bit-allocation coding mode in LPD is restricted to a lower bitrate at the encoder side. Therefore, to extend LPD-based stereo-coding mode, the bit allocation of LPD mode should first be revised. The ACELP mode comes originally from the AMR-WB coder, and the possible bit allocation range in the ACELP codebook is hence maximally restricted to 64×4 bits. We extended the two bit-allocation tables to 72×4 bits and 88×4 bits, which successfully supported stereo-coding mode up to 64-kb/s operation. Although the extended ACELP coding mode shows effective performance for speech items up to 64 kb/s, when the bitrate is above this value, MDCT-based LPD is preferably selected, according to the closed-loop based decision. Regarding the stereo-coding tool, USAC basically uses MPEG surround 2-1-2 mode (MPS212) for stereo coding [7], [14]. All coding modes in MPS212 can be successfully connected to the proposed LPD extension mode. For the residual coding mode of MPS212, however, bit-allocation logic can be considered, because the MPS212 outputs—the middle and residual signals—are actively estimated using spatial parameters, and have different dynamic ranges. Therefore, the bit-allocation logic used to encode a middle signal is set to a relatively higher level based on a comparison of the dynamic range and signal entropy between the middle and residual signals. All the revisions deal with encoder issues, and therefore the decoder does not need to be changed.

V. Performance Evaluation

1. Subjective Evaluation

Given that we have previously evaluated the proposed LPD mode extension for low-bitrate mono in [11], a subjective listening test was carried out in the present study for high-bitrate stereo at 64 kb/s; the additional extended elements of this work were also evaluated. The test items used here were constituted by four speech signals, four music signals, and three mixed speech-music signals, all of which were officially selected by the MPEG Audio group during the development of USAC, as shown in Table 1 [13]. The method for the subjective assessment of intermediate sound quality (MUSHRA) methodology [14]—which is commonly used in evaluations by the MPEG Audio group—was adopted as our evaluation method. Eleven expert listeners—who have experienced this type of tests several times—participated in our experiment.

Three systems—USAC (usac), the proposed extension for LPD-based USAC (elpd), and an elpd benchmark system (belpd)—were evaluated during the test, as shown in Table 2. A hidden reference (org) and a 3.5-kHz band-limited anchor (lp35) were added, according to the MUSHRA methodology. The benchmark system (belpd), which was developed in our

Table 1. Test items.

Test items	Category	Description
Speech1	Speech	English speech
Speech2		French speech
Speech3		German speech
Speech4		Korean speech
Music1	Music	Classical chorus music
Music2		Classical music
Music3		Rock music
Music4		Pop music
Mixed1	Mixed	Speech with chorus music
Mixed2		Speech with pop music
Mixed3		Speech with background music

Table 2. Description of systems evaluated under subjective testing.

System index	System description
elpd	Proposed extension of LPD-based USAC system
belpd	Benchmark system based on previous work [6]
usac	USAC system with the highest original quality
org	Hidden reference
lp35	3.5 low-pass filtered hidden anchor

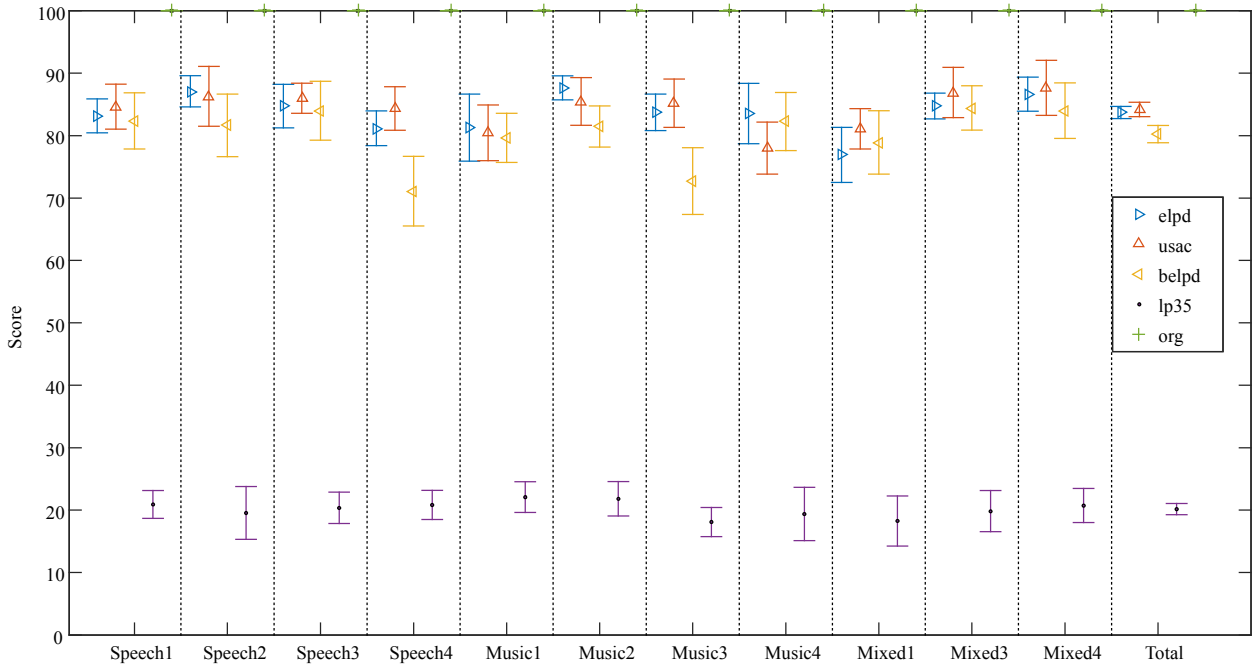


Fig. 7. Statistical analysis of the subjective evaluation: average absolute MUSHRA scores for 64-kb/s stereo.

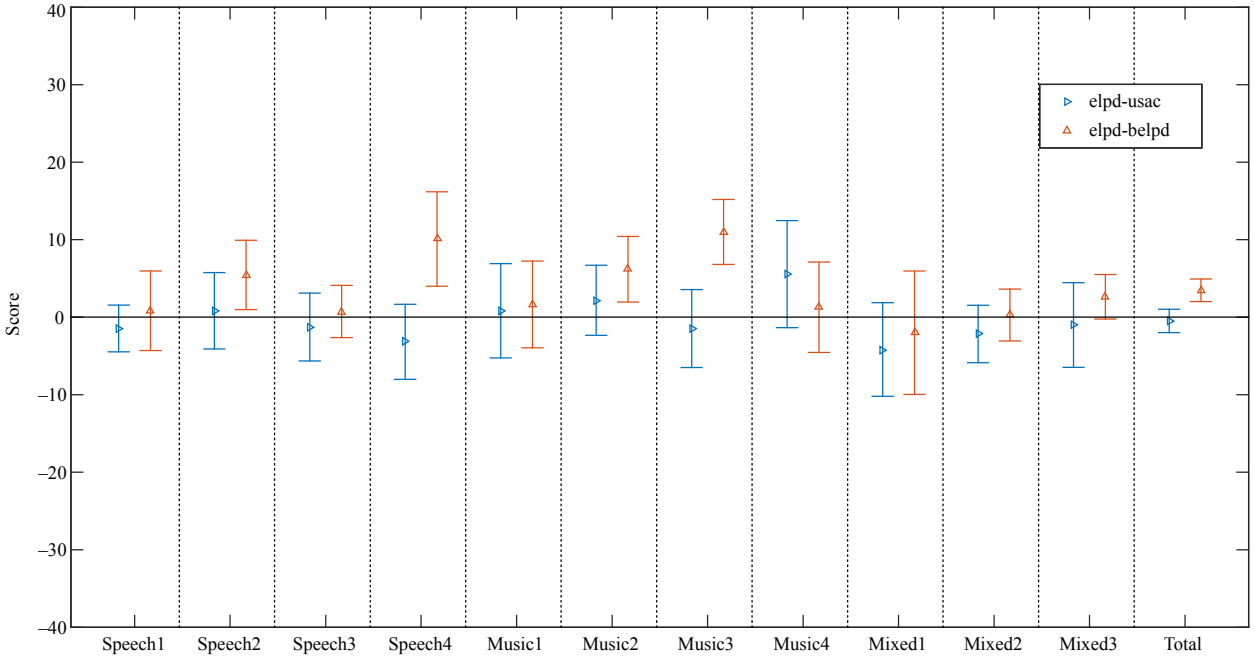


Fig. 8. Statistical analysis of the subjective evaluation: average difference in MUSHRA scores for 64-kb/s stereo.

previous study and shows a level of performance comparable with that of the original switching-based reference-quality USAC system, is limited to low-bitrate mode [6]. In our evaluation, we checked whether the proposed LPD-based USAC works effectively even at a higher bitrate, and confirmed its improvements when compared with the benchmark system. All the USAC bitstreams used during the tests are original reference-quality bitstreams and have demonstrated the highest level of quality offered by MPEG.

Figure 7 shows the average absolute score of the 64-kb/s stereo case, with a 95% confidence interval. The elpd and usac systems demonstrated a statistically overlapping performance for all items, and the overall average absolute score shows that the two systems are statistically equivalent within this 95% confidence interval. However, the overall average absolute score of the benchmark system, belpd, is statistically worse

https://doi.org/10.4218/etrij.17.0116.0397

than that of elpd, indicating the effectiveness of our proposed extension. A more detailed comparison can be obtained by considering the difference scores shown in Fig. 8. In the difference comparison, when the lower bound of the confidence interval does not include the zero value, we can state that the proposed scheme is better than the comparison systems in a statistically significant way. The two systems (elpd and usac) show the same level of statistical performance, both for the average difference scores of all items and for the overall score. In addition, an improvement from our previous studies is clearly shown in the difference scores between elpd and belpd.

2. Complexity

In this section, the structural complexity of the proposed USAC scheme is discussed, based on a comparison with the original USAC switching system. Our proposed coding components can create a USAC with a single mode of operation (LPD mode). This means that the additional coding tool regarding the switching system can be disabled. This disabled coding tool can lead to a reduction in structural complexity; instead of implementing the entire USAC system, practical implementations can be easier to achieve.

Table 3 shows the coding components that may be disabled when using the LPD mode extension; the signal classification to decide the coding mode (FD or LPD) is not necessary in the proposed single mode system. Normally, the signal classifier needs to look at the information from several thousand samples in advance to choose the coding mode of the next frame, thereby resulting in an additional coding delay at the decoder side; however, the single mode system is not affected by the delay compensation produced by signal classification. In the proposed single mode approach, the other USAC compensation tools applicable when switching between FD and LPD are not necessary Most of the complicated tools in USAC come from transition compensation, such as the forward aliasing cancellation tool used in MDCT processing, and the transition windowing handling between FD and LPD. Our proposed single mode system only requires TD tools, while maintaining a level of performance comparable to a coding scheme without a switching process.

Table 3. List of components that may by disabled, when compared with the original USAC system.

Coding components	Original USAC system	Proposed single mode system
Signal classifier	○	×
Delay compensation of classifier	○	×
Transition compensation	○	×
FD	○	×
TD	○	○

VI. Conclusion

Significant performance improvements in speech and audio coding were achieved by introducing switchable-coding based USAC. This technology plays a pivotal role as a bridge between audio and speech coding, and exhibits performances superior to those of existing state-of-the-art coding technologies such as HE-AACv2 and AMR-WB+. In this paper, we confirmed that the proposed single-mode operation USAC still works well compared to the original USAC, even for stereo at a higher bitrate. A revision of the current mode-decision mechanism in LPD can be applied to select the optimal coding mode when considering both the intra- and inter-superframe characteristics with a marginal loss in terms of the coding delay. This is left for further consideration.

References

[1] A. Gersho, "Advances in Speech and Audio Compression," *Proc. IEEE*, vol. 82, no. 6, June 1994, pp. 900–918.

[2] ISO/IEC JTC1/SC29/WG11, *Unified Speech and Audio Coding Verification Test Report*, Torino, Italy, MPEG 2011/N12232, July 2011.

[3] J.D. Johnston, "Transform Coding of Audio Signals Using Perceptual Noise Criteria," *IEEE J. Sel. Areas Commun.*, vol. 6, no. 2, Feb. 1988, pp. 314–323.

[4] 3GPP, *Adaptive Multi-Rate - Wideband (AMRWB) Speech Codec; General Description*, 3GPP TS 26.171, 2002.

[5] B. Bessette et al., "The Adaptive Multirate Wideband Speech Codec (AMRWB)," *IEEE Trans. Speech Audio Process.*, vol. 10, no. 8, Nov. 2002, pp. 620–636.

[6] J. Makinen et al., "AMR-WB+: a New Audio Coding Standard for 3rd Generation Mobile Audio Services," *IEEE Int. Conf. Acoustics Speech Signal Process.*, Philadelphia, PA, USA, Mar. 23, 2005, pp. 1109–1112.

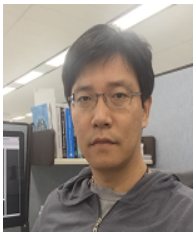
[7] M. Neuendorf et al., "The ISO/MPEG Unified Speech and Audio Coding Standard: Consistent High Quality for All Content Types and at All Bit Rates," *J. AES*, vol. 61, no. 12, Dec. 2013, pp. 956–977.

[8] S. Quackenbush, "MPEG Unified Speech and Audio Coding," *IEEE Multimedia*, vol. 20, no. 2, Apr.–June 2013, pp. 72–78.

[9] R.M. Aarts and R.T. Dekkers, "A Real-Time Speech-Music Discriminator," *J. Audio Eng. Soc.*, vol. 47, no. 9, Sept. 1999, pp. 720–725.

[10] J.G.A. Barbedo and A. Lopes, "A Robust and Computationally Efficient Speech/Music Discriminator," *J. Audio Eng. Soc.*, vol. 54, no. 7–8, July 2006, pp. 571–588.

[11] T. Lee et al., "Adaptive TCX Windowing Technology for Unified Structure MPEG-D USAC," *ETRI J.*, vol. 34, no. 3, June 2012, pp. 474–477.
[12] ISO/IEC 23003-3:2012, *MPEG-D (MPEG audio technologies), Part 3: Unified Speech and Audio Coding*, 2012.
[13] ISO/IEC SC29 WG11 N9638, *Evaluation Guidelines for Unified Speech and Audio Proposals*, MPEG, Jan. 2008.
[14] International Telecommunication Union, *Method for the Subjective Assessment of Intermediate Sound Quality (MUSHRA)*, ITU-R, Recommendation BS, 1543-1, Geneva, Switzerland, 2001.



Seungkwon Beack received his BS degree in electronic engineering from Korea Aviation University, Goyang, Rep. of Korea, in 1999, and his MS and PhD degrees in the Department of Information and Communications Engineering at the Korea Advanced Institute of Science and Technology, Daejeon, Rep. of Korea, in 2001 and 2005, respectively. He is currently with the ETRI, Daejeon, Rep. of Korea. His research interests include audio signal processing, and multi-channel audio coding and representation.

Jongmo Seong received his BS and MS degrees in electronics engineering from Pusan National University, Rep. of Korea, in 1995 and 1997, respectively. He received his PhD degree in mechatronics engineering from Chungnam National University, Daejeon, Rep. of Korea, in 2014. Since 1999, he has been working as a principal researcher in the Realistic AV Research Group at the ETRI, Daejeon, Rep. of Korea. His research interests cover a wide range of topics in speech and audio signal processing.

Misuk Lee received her BS and MS degrees in electronics engineering from Hoseo University, Asan, Rep. of Korea, in 1991 and 1993, respectively, and her PhD degree in electrical and electronics engineering from the Korea Advanced Institute of Science and Technology, Daejeon, Rep. of Korea, in 2001. Since February 2002, she has been with the ETRI, Daejeon, Rep. of Korea. Her current research interests include digital speech and audio coding, and digital audio signal processing techniques for immersive broadcasting.

Taejin Lee received his BS and MS degrees in electronics engineering from Chonbuk National University, Jeonju, Rep. of Korea, in 1996 and 1998, respectively, and his PhD degree in electronics engineering from Chungnam National University, Daejeon, Rep. of Korea, in 2013. He worked for Mobens Co., Ltd., Daejeon, Rep. of Korea, from 1998 to 2000. He has been with the ETRI since 2000, and he is now a principal researcher of the Realistic AV Research Group. From 2002 to 2003, he was a visiting researcher at Tokyo Denki University, Japan. His research interests include audio signal processing and interactive broadcasting technologies.