❏   146

Performance Evaluation of Multichannel Audio Compression

Teddy Surya Gunawan[1], Mira Kartiwi[2]
[1] Department of Electrical and Computer Engineering, Kulliyyah of Engineering, India
[2] Department of Information Systems, Kulliyyah of ICT
International Islamic University Malaysia

Article Info	ABSTRACT
	In recent years, multichannel audio systems are widely used in modern sound devices as it can provide more realistic and engaging experience to the listener. This paper focuses on the performance evaluation of three lossy, i.e. AAC, Ogg Vorbis, and Opus, and three lossless compression, i.e. FLAC, TrueAudio, and WavPack, for multichannel audio signals, including stereo, 5.1 and 7.1 channels. Experiments were conducted on the same three audio files but with different channel configurations. The performance of each encoder was evaluated based on its encoding time (averaged over 100 times), data reduction, and audio quality. Usually, there is always a trade-off between the three metrics. To simplify the evaluation, a new integrated performance metric was proposed that combines all the three performance metrics. Using the new measure, FLAC was found to be the best lossless compression, while Ogg Vorbis and Opus were found to be the best for lossy compression depends on the channel configuration. This result could be used in determining the proper audio format for multichannel audio systems.

Corresponding Author:

Teddy Surya Gunawan,
Department of Electrical and Computer Engineering, Kulliyyah of Engineering,
International Islamic University Malaysia,
Jalan Gombak, 53100 Kuala Lumpur, (+603) 6196 4521.
Email: tsgunawan@iium.edu.my, tsgunawan@gmail.com

1. INTRODUCTION

In recent years, multichannel audio systems are widely used in modern sound devices. Usually, two digits separated by a decimal point, e.g. 2.1, 4.1, 5.1, 6.1, 7.1, are used to classify the various kinds of speaker set-up [1], [2]. This number represents the number of audio tracks used. Some audio systems only have a single channel or two channels (stereophonic sound or 2.0 channel sound). The first digit shows the number of primary channels, i.e. satellite units, each of which are reproduced on a single speaker which has the capability to handle range of frequency between 100Hz to 22 kHz. On the other hand, the second digit (decimal digit) represents the presence of LFE (Low Frequency Effect) that is reproduced on a subwoofer. Moreover, surround system describes a type of audio output in which the sound appears to surround the listener by 360 degrees, in which it gives impression that sound are coming from all possible directions. It has been used to provide a more realistic and engaging experience [3].

There are two kinds of audio compression algorithm those are lossy and lossless. Lossy audio compression is known by their well-designed system to shrinks file sizes. Advanced Audio Coding (AAC), MPEG-1 Layer III (MP3), Dolby AC-3, Opus, OGG Vorbis [4] and Windows Media Audio Lossy (WMA lossy) are the examples of popular lossy audio coding system [5]. AAC can be considered as the most influential multichannel audio coding algorithm [6]. This is due to its ability to support audio channels up to 48 channels and contribute lossless audio for 5.1 channels at sampling rates 320 kbits/s. Meanwhile, AC3 provides high audio quality at 384kbit/s [7]. Meanwhile, the most well-known codec in lossless algorithm are Free Lossless Audio Codec (FLAC), Apple Lossless Audio Codec (ALAC), WavPack (WAV), MPEG-4

Audio lossless [8], True Audio (TTA) [9], and IEEE 1857.2 [10]. Lossless compression algorithms do not have any loss information and provide an exact replica of the original signal.

Although many research has been conducted on lossless and lossy audio compression, but not many researches have been focused on the performance evaluation on multichannel audio coding. Therefore, the objective of this paper is to investigate the performance of various audio compression algorithms to encode multichannel audio in terms of encoding time, data saving, and quality. Furthermore, a new integrated metric was proposed to integrate all three metrics.

2.   MULTICHANNEL AUDIO CONFIGURATION

The details of multichannel audio speaker configuration has been presented in [11]. From analog audio, sampling and quantization are conducted to represent the sound wave into digital representation. A stereo signal can be considered as two independent channels of audio information, i.e. left and right channels. Stereophonic audio provides the impression of sound localization. Unlike mono and stereo audio, multi-channel audio format designates in more than two channels. This type of audio format aims to advance the ability of sound localization. As an example, a 5.1 multichannel loudspeakers arrangement has been illustrated in Figure 1(a). The left and right channels placed at ±30˚ like in stereo audio. Meanwhile, the rear right and left channel located at ±110˚. Usually, they are used for extended sound source localizations interpretation. For center channel, 0˚ commonly for playing again voice contents in moving audio. The decimal digit (.1) channel refer to subwoofer channel which also recognize as LFE channel. This channel is for playing back the low frequency contents. By adding more surround loudspeaker to the two standard channels LS and RS, it will create larger listening zone. This setup had been widely used in cinema [12].
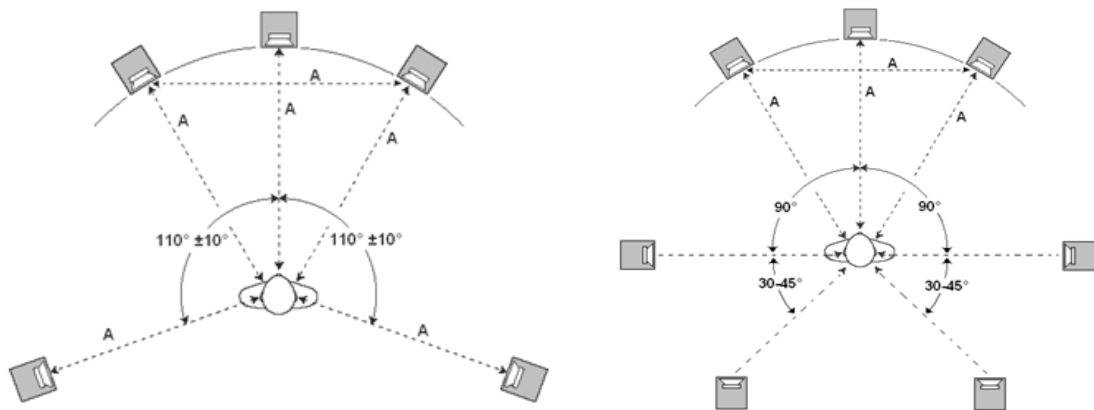

Figure 1. 5.1 and 7.1 Multichannel Speakers Setup

Multichannel audio 7.1 is a further enhancement to 5.1 audio channels. There are other two side-surround speaker in the speaker configuration. Many of application used 7.1 audio in order to greater impact of surround sound. The loudspeakers arrangement is almost similar to 5.1 multichannel audio. However, there are another two speaker left and right rear which about ±135˚to surround sound. Figure 1(b) shows the setup configuration of multichannel 7.1 audio, while Table 1 shows the standard channel layouts for multichannel audio. Beyond 7.1 multichannel audio, 10.2 channel surround sound has been developed. It is the advanced version of 5.1 technology, but 10.2 could produce twice as good as 5.1. In this channel configuration. 14 channels are used to including five front speakers, five surround channels, two LFE and two heights, plus the addition of a second sub-woofer [12].

Table 1. Standard Channel Layouts

Channel Name	Decomposition
Mono	FC
Stereo	FL + FR
2.1	FL + FR + LFE
5.1	FL + FR + FC + LFE + BL + BR
7.1	FL + FR + FC + LFE + BL + BR + SL + SR

FL = Front Left, FR = Front Right, FC = Front Center, LFE = Low Frequency
BL = Back Left, BR = Back Right, SL = Side Left, SR = Side Right

3.   MULTICHANNEL AUDIO COMPRESSION ALGORITHMS

In this paper, three lossy and three lossless audio compression algorithms will be evaluated, including Advanced Audio Coding (AAC), Ogg Vorbis, Opus, FLAC, TrueAudio, and WavPack, respectively. Note that, the selected coders are capable to handle multichannel compression for stereo, 5.1, and 7.1 channels.

3.1. Advanced Audio Coding (AAC)

AAC leads MP3 as there is a new non-backward compatible audio coder introduced in [1], [6]. It becomes popular due to application in Apple iTunes. AAC operates MDCT transform only in its main coding loop and transient detection function to detect a long window of 2048 points or a serial set of eight 256 point windows is ready for the MDCT transform. Thus, this give high frequency resolution of 23Hz and 2.7ms for a signal sampled at 48 kHz. A gain control procedure is incorporated in the SSR profile of AAC. A Pseudo Quadrature Mirror Filter (PQMF) filter bank is used to split the signal into four subbands with same bandwidth. The original signal sampling rates reduced to quarters by discarding one or more subbands. AAC utilizes the temporal-noise-sharping technique to expel the pre-echo effect caused by transients. Based on subjective evaluations, AAC provides great audio for 5 channel bandwidth at bit rate of 320kbps.

3.2. Ogg Vorbis

Ogg Vorbis is a full open source, non-proprietary, patent and royalty free compression audio format. It is based on vector quantization and transformation with overlapping windows, i.e. modified discrete cosine transform (MDCT). Each windows can have 2048 or 512 samples. The shorter one is used only to encode a transient signals. After transformation to frequency domain, the signal is analyzed by psychoacoustic model and inaudible part of the spectrum is removed. Then the floor vector is generated for each of the channels.

3.3. Opus

Similar with Ogg Vorbis, Opus is a full open source, non-proprietary, patent and royalty free compression audio format. It is standardized by the Internet Engineering Task Force (IETF) as RFC 6716 in September 2012. It is designed for a wide range of applications and scales from low bitrate narrowband speech at 6 kbit/s to very high quality stereo music at 510 kbit/s. Almost similar with other audio coder, it uses linear prediction and MDCT. The Opus format has three different modes, including speech, hybrid, and constrained energy lapper transform (CELT). The basic speech mode is using SILK algorithm developed by Skype mainly for speech signal, while CELT was used mainly for general audio signals. The hybrid mode uses SILK for the speech and uses CELT for the frequency range above 8000 Hz.

3.4. Free Lossless Audio Coding (FLAC)

Free Lossless audio coding (FLAC) is on of the most popular lossless codec due its fastest decoding audio. FLAC uses a linear prediction (LP) operation where future values of the digital signal are estimated as a linear function of previous samples. The FLAC encoder first divide the input audio signal into frames. Then, it will conduct an interchannel decorrelation. The predictor is utilized to find an optimum coefficients to predict the signal. Lastly, the predictor coefficients and its residue were passed to entropy coding.

3.5. TrueAudio (TTA)

TrueAudio is a free, open source, and real time lossless audio compressor for multichannel 8, 16, and 24 bits audio data, with the ability of password based data protection. It was designed to have reasonable compression levels while maintaining high operation speeds. The compression ratio can achieve as much as 30% of original file size, while it has real time encoding algorithm.

3.6. WavPack (Wv)

WavPack is another free and open source lossless audio compression algorithm. In the default lossless mode, WavPack acts just like a WinZip compressor for audio file with compression ratio between 30% to 70% depends on the audio source. The hybrid mode provides a relatively small, high quality lossy file that can be used all by itself, and a correction file that provides full lossless restoration. WavPack employs well known algorithms, such as linear prediction with least-mean-squared (LMS) adaptation, Elias and Golomb codes for entropy coding.

4.   RESULTS AND DISCUSSION

This section will discuss the audio database preparation, experimental setup, implementation, performance metrics, as well as performance evaluation.

4.1. Experimental Setup, Implementation and Audio Database

A high performance system was used for processing, i.e. a multicore system with Intel Core i7 6700 K 4.00 GHz (4 cores with 8 threads), 32 GBytes RAM, 256 GBytes SSD and 2 TBytes hard disk, installed with Windows 10 operating system and Matlab 2017b with Signal Processing Toolbox. To minimize the effect of other applications to the simulation, Windows 10 was booted in Safe Mode, in which Matlab was running with no java virtual machine, i.e. Matlab –nojvm. Similar to [11], the latest FFmpeg version 3.4.1 was used for the implementation of three lossy and three lossless audio coders. Matlab system call dos() was used to call FFmpeg executable.

The audio database was extracted from Ambra Experience Album (2008) which has DTS 5.1 (44.1 kHz, 16 bits) and FLAC 7.1 (48 kHz, 16 bits) format. The stereo signals were downmixed from the 5.1 audio source. Out of 10 tracks, we randomly selected three tracks for our experiments as shown in Table 2. Note that, the 7.1 channels has bigger file size due to its higher sampling frequency and eight channels in total.

Table 2. Audio Database for Multichannel Audio

Audio	Track	Length	File Size (in MBytes)		
			2	5.1	7.1
Audio1	01 Ambra – Honour and Glory	2 minutes and 22 seconds	23.8	71.5	868
Audio2	03 Ambra – Prism of Live	3 minutes and 28 seconds	35	105	152
Audio3	06 Faszination Natur – Seven Seasons	3 minutes and 3 seconds	30.8	92.5	134

4.2. Performance Measures

To evaluate the performance of audio coders, encoding time and percentage data reduction were measured for each coder and each audio file. For encoding time (E_T) accuracy, the Matlab program will loop 100 times ($N = 100$) and the average value is taken as shown in Eq. (1). The percentage data reduction is measured as shown in Eq. (2).

$$E_T = \frac{1}{N} \sum_{n=1}^{N} T_n \tag{1}$$

$$D_R = \left(1 - \frac{B_{original}}{B_{encoded}}\right) \times 100\% \tag{2}$$

where $B_{orginal}$ is the original file size in bytes and $B_{encoded}$ is the encoded file size in bytes.

For lossless audio compression, such as FLAC, TrueAudio, and WavPack, there is no loss in audio quality. But for lossy compression, such as AAC, Ogg Vorbis, and Opus, there will be loss in audio quality which can be measured subjectively using listening test or objectively using PEAQ [13]. PEAQ has been standardized as ITU-R BS.1387-1 has two main parts, which is the psychoacoustic model and the cognitive model. Up until now, PEAQ is only able to measure the objective difference grade (ODG) for up to stereo signals. In [14], the authors proposed the extension of PEAQ for multichannel audio. However, it has not been adopted as new standard yet until now. The ODG score can range from 0 to -4, in which 0 represents a signal with imperceptible distortion and -4 represents a signal with very annoying distortion. Furthermore, in this paper the advanced version of PEAQ which has two peripheral ear models and filter bank based ear models was used due to its accuracy. Note that, the current PEAQ limitation, the PEAQ measurement will be conducted only on stereo signals and denoted as Q as shown in Eq. (3).

$$Q = \Psi(x, \hat{x}) \tag{3}$$

where is Ψ is the PEAQ function, x is the original WAV file, and $\hat{x}$ is the encoded-then-decoded WAV file.

4.3. Time and Frequency Spectrum of Multichannel Audio Signal

Figure 2 shows the example of time domain and frequency spectrum of 7.1 audio signal (Audio1). From this figure, it can be seen that interchannel decorrelation could be conducted between left and right channel (front, back, and side), i.e. mid and side signals, while front center and LFE could be encoded separately as practiced by many multichannel audio coders.

(a) Front Left channel

(b) Front Right channel

(c) Front Center channel

(d) Low Frequency channel

(e) Back Left channel

(f) Back Right channel
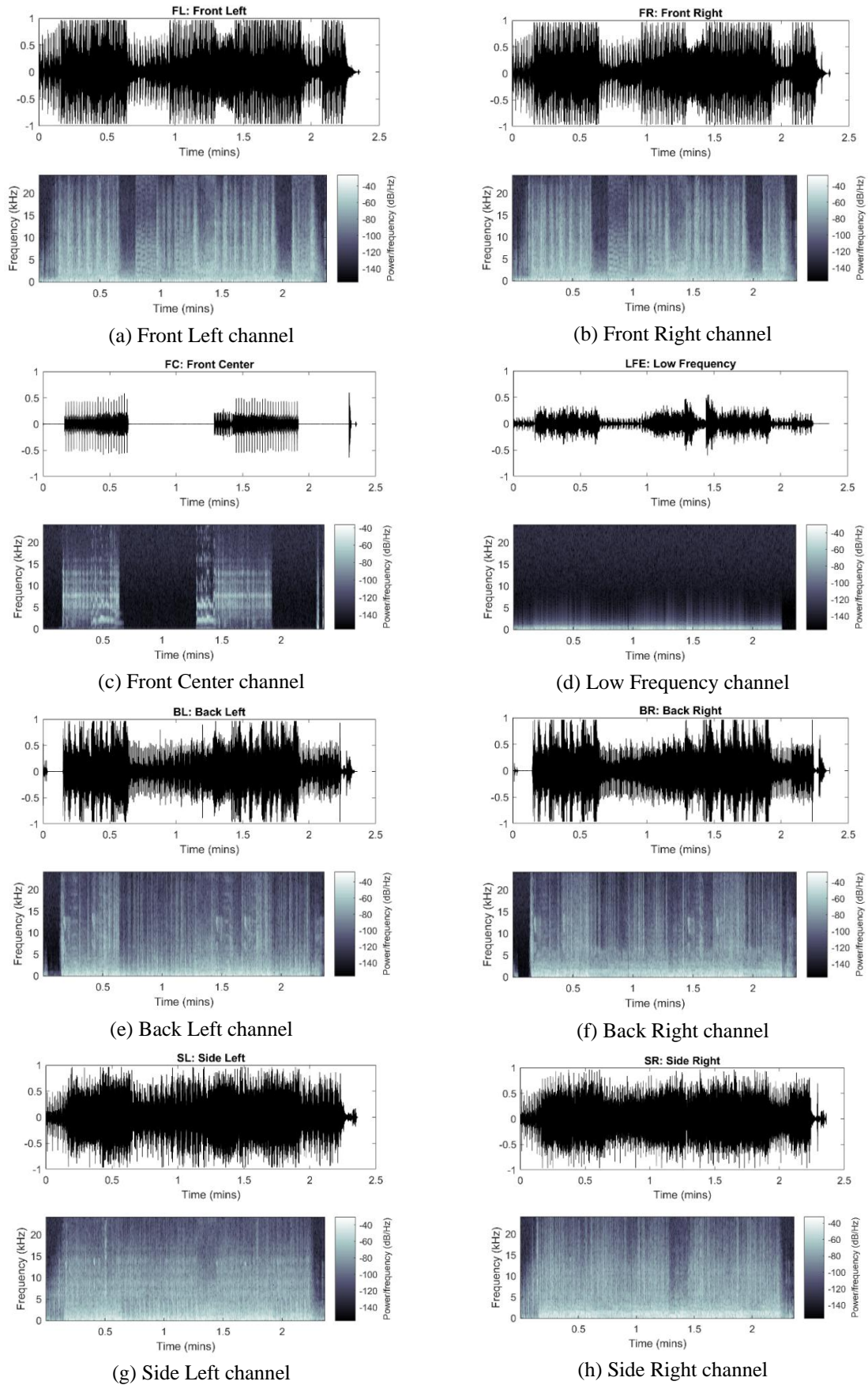

(g) Side Left channel

(h) Side Right channel

Figure 2. Sample of time domain and frequency spectrum of 7.1 audio signal

4.4. Experiments on Lossy and Lossless Compression

Table 3 and 4 shows the data reduction (%) and average encoding time (seconds) for lossy and lossless compression of stereo, 5.1, and 7.1 audio signals using AAC, Ogg, Opus, FLAC, TrueAudio, and WavPack. Across various channel configuration, the average data reduction for lossy compression is 91.20%, 92.31%, and 92.86% for AAC, Ogg Vorbis, and Opus. Opus has the highest compression compared to the other algorithms. Meanwhile, the average data reduction for lossless compression is 51.63%, 47.23%, 48.93% for FLAC, TrueAudio, and WavPack, respectively. It has been found that FLAC has the highest compression compared to the other algorithms.

From Table 3 and Table 4, across various channel configurations, the average encoding time for lossy compression is 8.19 seconds, 5.69 seconds, and 9.46 seconds for AAC, Ogg Vorbis, and Opus, respectively. Meanwhile, the average encoding time for lossless compression is 1.89 seconds, 2.16 seconds, and 2.11 seconds for FLAC, TrueAudio, and WavPack, respectively. For lossy compression of stereo signals, another performance could be measured which is the quality, i.e. PEAQ ODG. Based on Table 5, Ogg Vorbis has the highest quality with big margin compared to AAC and Opus.

Table 3. Data Reduction (%) for Various Encoders

Encoder	Channel	Audio1	Audio2	Audio3	Average
	2	90.57	90.53	90.8	90.63
AAC	5.1	90.6	90.6	90.72	90.64
	7.1	92.29	92.3	92.41	92.33
	2	92.28	92.38	92.75	92.47
Ogg Vorbis	5.1	92.9	93.23	94.19	93.44
	7.1	90.78	90.98	91.33	91.03
	2	92.48	92.8	93.12	92.80
Opus	5.1	92.51	93.01	92.37	92.63
	71	92.8	93.32	93.36	93.16
	2	49.19	47.69	60.58	52.49
FLAC	5.1	51.29	52.65	58.49	54.14
	7.1	46.2	47.28	51.34	48.27
	2	50.17	48.44	61.11	53.24
TrueAudio	5.1	41.93	43.41	52.58	45.97
	7.1	39.01	40.58	47.87	42.49
	2	46.36	44.81	56.77	49.31
WavPack	5.1	48.74	50.82	56.15	51.90
	7.1	43.44	45.18	48.11	45.58

Table 4. Encoding Time (seconds) for Various Encoders

Encoders	Channel	Audio1	Audio2	Audio3	Average
	2	2.9	4.39	4.39	3.42
AAC	5.1	8.94	12.62	12.62	9.82
	7.1	9.91	14.14	14.14	11.32
	2	1.88	2.64	2.64	2.29
Ogg Vorbis	5.1	5.33	7.56	7.56	6.39
	7.1	6.81	9.85	9.85	8.40
	2	1.29	1.88	1.88	1.76
Opus	5.1	3.52	5.05	5.05	4.68
	71	4.21	6.29	6.29	21.95
	2	0.36	0.48	0.48	0.83
FLAC	5.1	0.75	1.11	1.11	2.02
	7.1	1.11	1.55	1.55	2.83
	2	0.42	0.57	0.57	0.89
TrueAudio	5.1	1.11	1.55	1.55	2.33
	7.1	1.55	2.21	2.21	3.27
	2	0.4	0.54	0.54	0.87
WavPack	5.1	0.99	1.44	1.44	2.24
	7.1	1.55	2.11	2.11	3.22

Table 5. Objective Quality for Various Encoders on Stereo Signals

Encoder	Objective Difference Grade (ODG)			
	Audio1	Audio2	Audio3	Average
AAC	-3.38	-3.49	-3.38	-3.42
Ogg Vorbis	-0.66	-0.76	-0.63	-0.68
Opus	-3.58	-3.58	-3.56	-3.57

4.5. New Integrated Performance Metric

Based on previous discussion, it is rather difficult to evaluate the performance of each encoder as it needs to evaluate at least two metrics at the same time, i.e. encoding time and data saving, and quality as well in the case of lossy compression for stereo signals. It is well known that there always will be a trade-off between encoding time (complexity) and data saving. The integrated measurement metric should be taking care all of measurement metrics. We proposed the following new integrated performance metric:

$$\vartheta = f(E_T, D_R, [Q]) = \gamma \times \frac{D_R}{E_T \times |Q|} \qquad (4)$$

where E_T is the encoding time (in seconds), D_R is the data reduction or saving (in %), and Q is the quality (ODG value) which is only applicable for lossy compression (and up to stereo signals at the moment), and γ is a measurement constant (in seconds). The integrated metric was derived from the following reasons. The performance of an audio encoder is proportional to the data reduction, D_R, and inversely proportional to E_T and ODG $|Q|$. For the lossless compression, the ϑ will be depends only on D_R and E_T, and Q could be set to a very small number (representing an exact replica and no loss in information). For our implementation, let us set $\gamma = 1$ for both lossless and lossy compression, and $Q = 0.1$ for lossless compression (to represent high quality).

From Table 6, now we can evaluate the performance of each encoder in one performance measure, ϑ, per channel configuration. The best performance between channel and between lossless and lossy compression were highlighted in bold. Among the lossy encoders for stereo signals, Ogg Vorbis has the highest performance. However, Opus has the highest performance for 5.1 and 7.1 channel. Among the lossless compression, FLAC has the highest performance for all channel configuration. In conclusion, our integrated measure, ϑ, is able to capture the performance of each encoder in terms of encoding time, data saving, and quality.

Table 6. Performance Evaluation of Various Encoders using Integrated Metric ϑ

Encoder	Files	Channel		
		2	5.1	7.1
AAC	Audio1	9.2	10.1	9.3
	Audio2	5.9	7.2	6.5
	Audio3	6.1	7.2	6.5
	Average	*7.1*	*8.2*	*7.5*
Ogg Vorbis	Audio1	74.4	17.4	13.3
	Audio2	46.0	12.3	9.2
	Audio3	55.8	12.5	9.3
	Average	*58.7*	*14.1*	*10.6*
Opus	Audio1	20.0	26.3	22.0
	Audio2	13.8	18.4	14.8
	Audio3	13.9	18.3	14.8
	Average	*15.9*	*21.0*	*17.2*
FLAC	Audio1	1366.4	683.9	416.2
	Audio2	993.5	474.3	305.0
	Audio3	1262.1	526.9	331.2
	Average	*1207.3*	*561.7*	*350.8*
TrueAudio	Audio1	1194.5	377.7	251.7
	Audio2	849.8	280.1	183.6
	Audio3	1072.1	339.2	216.6
	Average	*1038.8*	*332.3*	*217.3*
WavPack	Audio1	1159.0	492.3	280.3
	Audio2	829.8	352.9	214.1
	Audio3	1051.3	389.9	228.0
	Average	*1013.4*	*411.7*	*240.8*

5.   CONCLUSIONS AND FUTURE WORKS

This paper has presented the performance evaluation of three lossy and three lossless compression for multichannel audio signals, including stereo, 5.1 and 7.1 channels. The six audio compression algorithm, i.e. AAC, Ogg Vorbis, Opus, FLAC, TrueAudio, and WavPack, have been confirmed to be able to perform compression up to 7.1 channel. Experiments were conducted on the same three audio files but with different channel configurations. The performance of each encoder was evaluated based on its encoding time (averaged over 100 times), data saving, and audio quality. Furthermore, we proposed one integrated performance measure ϑ to ease the evaluation. Using the new measure, FLAC was found to be the best

lossless compression, while Ogg Vorbis and Opus were found to be the best for lossy compression. Further research can be conducted including the use of different audio database, parameters optimization for each encoder, and the use of different audio coders.

ACKNOWLEDGEMENT

REFERENCES
[1] M. Bosi and R. E. Goldberg, "Introduction to digital audio coding and standards," Springer Science & Business Media, vol. 721, 2012.
[2] F. Rumsey, "Spatial audio," CRC Press, 2012.
[3] F. Schuh, *et al.*, "Efficient multichannel audio transform coding with low delay and complexity," in *Audio Engineering Society Convention*, pp. 141, 2016.
[4] J. Moffitt, "Ogg Vorbis—open, free audio—set your media free," *Linux journal*, vol. 2001, pp. 9, 2001.
[5] W. Jackson, "Audio Concepts, Terminology, and Codecs," in Android Apps for Absolute Beginners: Springer, pp. 651-663, 2014.
[6] M. Bosi, *et al.*, "ISO/IEC MPEG-2 advanced audio coding," *Journal of the Audio engineering society*, vol. 45, pp. 789-814, 1997.
[7] R. Hennequin, *et al.*, "Codec independent lossy audio compression detection," in *Acoustics, Speech and Signal Processing (ICASSP), 2017 IEEE International Conference on*, pp. 726-730, 2017.
[8] T. Liebchen, *et al.*, "The MPEG-4 Audio Lossless Coding (ALS) standard-technology and applications," in *Proc. 119th AES Conv*, 2005.
[9] A. Djuric, "TTA Lossless audio codec-True audio compressor algorithms," 2010.
[10] T. S. Gunawan, *et al.*, "Investigation of Lossless Audio Compression using IEEE 1857.2 Advanced Audio Coding," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 6, 2017.
[11] T. S. Gunawan, *et al.*, "Investigation of Various Algorithms on Multichannel Audio Compression," in *Proc. of the 4th IEEE International Conference on Smart Instrumentation, Measurement and Applications (ICSIMA)*, Putrajaya, pp. 1-5, 2017.
[12] T. Holman, "Surround sound: up and running," CRC Press, 2014.
[13] T. S. Gunawan, "Audio compression and speech enhancement using temporal masking models," Thesis, University of New South Wales, Sydney, Australia, 2007.
[14] J. Liebetrau, *et al.*, "Standardization of PEAQ-MC: Extension of ITU-R BS. 1387-1 to multichannel audio," in *AES 40th Int. Conf, Spatial Audio: Sense the Sound of Space*, Tokyo, Japan, 2010.