



## Performance evaluation of power-saving audio reproduction

Mitsunori Mizumachi, Tsukasa Nakashima, Mitsuhiro Nakagawara

### ► To cite this version:

Mitsunori Mizumachi, Tsukasa Nakashima, Mitsuhiro Nakagawara. Performance evaluation of power-saving audio reproduction. Forum Acusticum, Dec 2020, Lyon, France. pp.37-38, <10.48465/fa.2020.0321>. <hal-03235348>

**HAL Id: hal-03235348**

**<https://hal.science/hal-03235348v1>**

Submitted on 27 May 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

# PERFORMANCE EVALUATION OF POWER-SAVING AUDIO REPRODUCTION

Mitsunori Mizumachi<sup>1</sup> Tsukasa Nakashima<sup>1</sup> Mitsuhiro Nakagawara<sup>1,2</sup>

<sup>1</sup> Kyushu Institute of Technology, Kitakyushu, Japan

<sup>2</sup> Panasonic, Yokohama, Japan

## ABSTRACT

The trends of audio reproduction can be divided into three categories: high fidelity reproduction with high resolution for audio enthusiasts, immersive audio reproduction in the professional audio industry, and casual audio reproduction with portable audio players, smartphones, and tablets. Casual music listeners have concerned with the battery life of the mobile device while audio reproduction rather than the quality of audio reproduction. The authors have proposed a power-saving audio playback algorithm based on our auditory characteristics. In this paper, the relationship between perceptual evaluation and power consumption is investigated for both music and speech signals. For the music sources, general attributes used for audio quality assessment are employed to quantify the distortion caused by the proposed method. Concerning the speech signals, the diagnostic rhyme test is carried out to characterize the speech distortion when the reduction rates of power consumption are set at approximately 20 %, 50 %, and 70 %. It is confirmed that the proposed method could reduce the power consumption by 25 % and 50 % within the acceptable distortion for music and speech signals, respectively.

## 1. INTRODUCTION

Recent audio applications make more diverse on audio reproduction. The novel transducers in the high frequency domain reproduce high resolution audio sources. The use of multiple loudspeakers enables immersive audio reproduction. On the other hand, on-demand music streaming subscription services bring about a revolution in music listening style. A massive collection of music sources can be personally reproduced using smartphones and portable devices anytime and anywhere. In such cases, the listeners give priority to extending the battery life of the mobile devices rather than audio quality.

The authors have proposed a power-saving audio reproduction algorithm based on our static auditory characteristics [1] and the auditory masking [2]. Those algorithms aim at decreasing the power consumption with less perceptual distortion while reproducing audio sources. The feasibility of the proposed method was confirmed for some music sources.

In this paper, the relationship between perceptual evaluation and power consumption is investigated for both music and speech signals. The perceptual

degradation in some attributes used for audio quality assessment is measured for music signals by carrying out a listening test. The diagnostic rhyme test is carried out to characterize the speech distortion.

## 2. POWER-SAVING AUDIO REPRODUCTION ALGORITHM

The power-saving audio reproduction is achieved by controlling the sub-band power in the filterbank analysis. In this study, auditory-oriented filterbanks are introduced into sub-band signal processing. Three different filterbanks are considered such as the conventional octave filterbank, mel-frequency filterbanks with triangular windows, and the gammachirp filterbank, which is designed based on both physiological and psychophysical data [3,4]. The octave and mel-frequency filterbanks are not level-dependent, but the gammachirp filter broadens its main-lobe for the input with high intensity.

The proposed algorithm attenuates the low-frequency components up to 120 Hz in 3 dB, assuming that the lower components are mostly masked by background noises because the casual listeners are supposed to listen to music via earphones and headphones under noisy environments. In addition, 3 dB boost is applied in the mid-frequency range from 3 kHz to 4 kHz, where the human ears are more sensitive.

The power-saving audio reproduction algorithm achieves the further reduction of power consumption based on the concept of auditory masking. The psychoacoustic masking model [5] is implemented to calculate the masking threshold. The proposed method attenuates the sub-band components of which amplitude levels are below the masking threshold. Finally, both high-end and low-end frequency components are omitted because popular earphones and headphones cannot reproduce the components in high fidelity below 60 Hz and upper 15 kHz.

## 3. PERFORMANCE EVALUATION

The feasibility of the proposed method is subjectively confirmed both for music and speech signals.

### 3.1 Subjective evaluation of music signals

The performance of the proposed method was subjectively evaluated based on the degradation mean opinion scores (DMOSs) [6] with the five-grade scale. The attributes consist of the sense of sound pressure,

**Table 1.** DMOSs on sound quality evaluation for the proposed method with tree filterbanks and non-processed conditions.

	Octave filterbank	Gammachirp filterbank	Mel-frequency filterbank	Non process.
sense of sound pressure	<b>4.42</b>	4.04	4.08	4.64
clarity	4.32	4.20	<b>4.34</b>	4.70
perceptual distortion	<b>4.46</b>	4.28	4.28	4.72
resonance	4.16	4.06	<b>4.26</b>	4.68
comprehensive impression	<b>4.38</b>	4.00	4.14	4.72

clarity, perceptual distortion, resonance, and comprehensive impression. The genre of the sound sources included pops, rock, and violin solo. 50 students (12 females and 38 males) participated in the listening test, where the parameters of the proposed method are set to achieve the power reduction rate of 30% approximately.

The experimental results are given in Table 1. The proposed method does not cause fatal perceptual problems, although there is a slight difference in DMOS among the types of filterbanks.

### 3.2 Subjective evaluation of speech signals

Subjective evaluation was done by the diagnostic rhyme test with the Japanese word pairs [7]. 24 students (6 females and 18 males) participated in the diagnostic rhyme test. The parameters of the proposed method were set to achieve the power reduction rates of 22.2%, 48.6%, and 73.2%, respectively. As a reference, non-processed original word pairs were also prepared for the test.

The results of the diagnostic rhyme test are summarized in Table 2 for each of the phonetic features: voicing, nasality, sustention, sibilant, graveness, and compactness. The intelligibility scores maintain more than 80% even in the case that the power reduction rate is 73.2%.

## 4. CONCLUSIONS

The previously proposed power-saving audio reproduction algorithm was subjectively evaluated for music and speech signals. In music appreciation, there is no significant perceptual degradation when the reduction rate of power consumption is set to 30%. Concerning the speech applications, the proposed method could reduce the power consumption up to 70% for practical use.

Future works include performance evaluation in practical applications under real environments.

**Table 2.** Results of Japanese diagnostic rhyme test.

	Power reduction rate			
	73.2%	48.6%	22.2%	0%
Voicing	82.5	98.3	95.0	100.0
Nasality	92.9	97.1	97.5	97.5
Sustention	94.6	98.3	99.6	99.6
Sibilant	97.1	99.2	99.2	99.6
Graveness	84.2	91.3	90.4	95.4
Compactness	95.8	98.8	100.0	100.0

## REFERENCES

- [1] M. Mizumachi *et al.*, "Power Saving Audio Playback Algorithm Based on Auditory Characteristics," *Proc. 144<sup>th</sup> AES Convention*, Paper No. 9933, 2018.
- [2] T. Nakashima *et al.*, "Proposal of Power-Saving Audio Playback Algorithm Based on Auditory Masking," *Proc. 146<sup>th</sup> AES Convention*, Paper No. 10164, 2019.
- [3] T. Irino and R.D. Patterson, "A time-domain level-dependent auditory filter: the gamma- chirp," *J. Acoust. Soc. Am.*, 101 (1), pp. 412- 419, 1997.
- [4] T. Irino and R.D. Patterson, "A compressive gammachirp auditory filter for both physiological and psychophysical data," *J. Acoust. Soc. Am.*, 109(5), pp. 2008-2022, 2001.
- [5] D. O'Shaughnessy, *Speech communication: human and machine*, Addison-Wesley, 1987.
- [6] ITU-T Rec. P .800: "Methods for subjective determination of transmission quality," 1996.
- [7] M. Fujimori *et al.*, "On a revised word-pair list for the Japanese intelligibility test," *Proc. Intl. Symp. on Frontiers in Speech and Hearing Research*, 2006.