Application of Power-saving Playback Algorithm to Speech Signals

Mitsuhiro Nakagawara^{a,b,*}, Tsukasa Nakashima^b, Mitsunori Mizumachi^b

^aPanasonic Corporation, 4261 Ikonobe-cho, Tsuzuki-ku, Yokohama-shi, Kanagawa, 224-8520, Japan.

^bKyushu Institute of Technology, 1-1 Sensui-cho, Tobata-ku, Kitakyushu-shi, Fukuoka, 804-8550, Japan.

*Corresponding Author: nakagawara.mitsuhiro@jp.panasonic.com

Abstract

Power consumption is one of the important issues for portable electronic devices. The authors have previously proposed a power-saving audio playback algorithm based on auditory masking. In this paper, the feasibility of the power saving audio playback algorithm is confirmed towards speech signals in the viewpoints of both speech intelligibility and power consumption. The influence of the speech conversion for power saving playback was quantified as the word intelligibility. In practice, the word intelligibility was measured in each of six phoneme categories for the converted speech signals of which the reduction rate of power consumption was set at 20 %, 50 %, and 70 %. It is confirmed that the word intelligibility does not decrease drastically for the converted speech of which the reduction rate of power consumption was set at 50%.

Keywords: power saving playback, speech signal, word intelligibility.

1. Introduction

High fidelity audio focuses on high resolution and distortion-free reproduction, but for casual audio listeners using smartphones and smart speakers, power saving while audio playback is an important factor. Power-saving audio playback has been studied both in hardware and software implementation. In the hardware-based approaches, each functional circuit of audio playback systems needs to be redesigned. A power amplifier has been promoted the efficiency based on the class AB and class D topologies⁽¹⁾. A full digital audio system achieves energy saving by using digital loudspeakers with multiple coils⁽²⁾⁽³⁾. On the other hand, software attempts are relatively easier to be achieved by digital signal processing.

The authors approached a software-oriented for power-

saving audio playback based on auditory characteristics which are implemented with the filterbank process⁽⁴⁾ and the masking processing⁽⁵⁾. First, the proposed method focused on the musical source and achieved the reduction of power consumption by 25 %⁽⁴⁾. Then, the availability of the proposed method was confirmed using a variety of sound sources including speech signals in simulated noisy conditions. The results of the listening tests suggested that the sound quality was mostly satisfactory under noisy environments⁽⁶⁾. It is important how accurately we can understand meaningful words when we perceive speech⁽⁷⁾. In this paper, the relationship between the word intelligibility and the current consumption is investigated by carrying out the diagnostic rhyme test in Japanese.

2. Power-saving Audio Playback Algorithm

The overview of the proposed power-saving playback algorithm is summarized in figure 1⁽⁶⁾. The method mainly consists of the auditory-oriented sub-band optimization in the filterbank processing and the reduction of inaudible components below the masking thresholds.

It is supposed that the reduction ratio of current consumption affects the word intelligibility. Therefore, it is necessary to prepare some reduction ratios and investigate the relationship between both the reduction ratio and the word intelligibility. In this paper, the mel-frequency filterbanks are employed in the filterbank analysis shown in figure 1 because the mel-frequency filterbanks achieved the efficient power saving compared to other filterbanks⁽⁴⁾. The reduction ratios of current consumption are adjusted at approximately 20 %, 50 %, and 70 % by setting the number of filterbanks at 5, 15, and 25 when the whole frequency range of the filterbank is fixed up to 20 kHz, respectively.

The power-saving audio playback algorithm relies on the auditory masking for eliminating non-audible frequency

Fig. 1. Proposed power-saving playback algorithm.

components. The auditory masking occurs when the perception of the target sound is affected by the presence of another sound. The masking threshold is calculated by using the psycho-acoustic model used in the MPEG-1 codec⁽⁸⁾. Then, the frequency components below the masking threshold are eliminated before the filterbank analysis⁽⁵⁾.

After the filterbank analysis and synthesis, the sub-band power is increased up to 3 dB at the auditory sensitive frequency bands from 3 kHz to 4 kHz and decreased at 3 dB in power at 120 Hz and lower frequency components. Finally, band-pass filtering in between 40 Hz to 15 kHz was applied for cutting out the extra high and low frequency components.

3. Measurement of Power Consumption

Power consumption was measured using an electronic circuit shown in figure 2. The prepared speech signals were played back through a simplified power amplifier and a loudspeaker. The current consumption was obtained from the voltage in between the resistor, and it was averaged by measured five times. The speech signals were prepared from the male utterances recorded in a soundproof room.

Figure 3 shows the exact reduction ratios for the each versions of the filterbanks with 5, 15, and 25 channels, respectively.

Fig. 2. Measurement circuits of power consumption.

Fig. 3. The reduction ratio of current consumption with the number of filterbanks for proposed algorithm.

4. Diagnostic Rhyme Test

It is important for our speech communication to accurately convey the contents of the utterances rather than the sound quality. In general, speech intelligibility can be measured with phonemes, words, and sentences⁽⁹⁾. In this study, the word intelligibility is measured by carrying out the Diagnostic Rhyme Test (DRT) standardized by ANSI⁽¹⁰⁾. The six phoneme categories were used which defined by Jacobson et al as shown in Table 1. The Japanese DRT word pair lists shown in annex A.1.⁽¹¹⁾ were used in the experiment. In each word pair, the first phonemes are different and the latter phonemes are the same. The subjects listened to the above words, and were asked to write down what they have heard.

Subjects were 24 students, 6 females and 18 males, with normal hearing. The DRT was carried out in the soundproof room in Kyushu Institute of Technology. The listening tests was carried out as follows.

- (1) A word was randomly selected from the ten pairs of Japanese DRT words to be presented to the subject.
- (2) After listening the presented word, the subjects wrote down what they heard in the answer sheet shown in annex A.2 in three seconds.
- (3) The next word was automatically presented to the

Table 1. The phenome categories.

No.	Phenomes	Corresponding classification
1	Voicing	vocalic-nonvocalic
2	Nasality	nasal-oral
3	Sustention	continuant-interrupted
4	Sibilant	strident-mellow
5	Graveness	grave-acute
6	Compactness	compact-diffuse

subjects after the three seconds break. It continued until the whole words of the ten pairs were presented.

(4) The above steps: (1), (2), and (3) were repeated in each phoneme category.

Figure 4 shows the results of the DRT in each phoneme category for the original speech signals and the processed speech signals of which the power consumption reduction rates are 20 %, 50 %, and 70 %, respectively. It is confirmed that the word intelligibility rates over 80 % were obtained in all phoneme categories.

In particular, the word intelligibility rates are more than 90 % for the processed speech signals with the reduction rates of 20 % and 50 %. For more detailed discussion in each phoneme category, the word intelligibility is relatively low in the Graveness category. It might be due to the intelligibility of the recorded word utterances without any processing.

5. Conclusions

The power-saving audio playback algorithm is modified for speech reproduction systems. The number of mel-frequency filterbanks were adjusted to achieve the reduction of power consumption up to approximately 20 %, 50 %, and 70 %. The feasibility of the modified algorithm was confirmed by carrying out the Japanese Diagnostic Rhyme Test in six phoneme categories. The experimental results suggest that the word intelligibility rates attain 90 % and more in the whole phoneme categories for the processed speech signals with the reduction rates of 20 % and 50 %. Even for those with the reduction rate of 70 %, the intelligibility rate exceeds 80 %. The proposed method could drastically decrease the power consumption for speech signals compared to music sound sources. Future works include the performance tests under the practical conditions.

Fig. 4. Each Phenome with the reduction ration of current consumption.

References

- (1) P.Tournier, "Class AB versus D for Multimedia Applications in Portable Electronic Devices," AES 29th International Conference: Audio for Mobile and Handheld Devices, Paper No. 2-2, 2006
- (2) J.Okamura and A.Yasuda, "Digital speaker driving apparatus," U.S. Patent No.8, 306, 244, 6 Nov. 2012
- (3) Y.Furuya, M.Takahashi, S.Saikatsu, M.Yoshino and A.Yasuda, "Speaker system with 100-W high output power and 0.17% THD using 9-V power supply with digitally direct-driven technique," IEEE International Conference on Consumer Electronics, 2017

- (4) M.Mizumachi, W.Kubota and M.Nakagawara, "Power saving audio playback algorithm based on auditory characteristics," Proc. 144th AES Conv., 2018
- (5) T. Nakashima, M. Nakagawara and M. Mizumachi, "Proposal of power-saving audio playback algorithm based on auditory masking," Proc. 146th AES Conv., 2019
- (6) T. Nakashima, M. Nakagawara and M. Mizumachi, "Subjective evaluation of power saving audio playback algorithm based on auditory masking," IEICE Technical Report, 2019
- (7) Acoustical Society of Japan, "Introduction to Acoustics," Corona publishing, 2011
- (8) ISO/IEC 11172-3: Coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbit/s - Part 3: Audio, 1993
- (9) Acoustical Society of Japan, "Acoustic Keyword Book," Corona publishing, 2016
- (10) ANSI Standard S3.2-1989, Method for Measuring the Intelligibility of Speech over Communication Systems, (1989, reaffirmed 1995)
- (11) K.Kondo, R.Izumi, M.Fujimori, R.Kaga and K.Nakagawa, "On a Two-to-one Selection Based Japanese Speech Intelligibility Test," Journal of the Acoustical Society of Japan, 63(4), pp196-204, 2007

Annex

A.1 Japanese DRT words pair lists.

A.1 shows the Japanese DRT words pair lists.

A.2 Answer sheet of listening tests.

A.2 shows the answer sheet of listening tests.

A.1. The Japanese DRT words pair lists.

Voicing	Nasality	Sustention	Sibilant	Graveness	Compactness
財 - オ (ザイ - サイ)	万 - 番 (マン - パン)	箸 - 菓子 (ハシ - カシ)	ジャム - ガム (ジャム - ガム)	杵 - 楽 (ワク - ラク)	焼く - 沸く (ヤク - ワク)
抱く - 炊く (ダク - タク)	無い - 台 (ナイ - ダイ)	旗 - 型 (ハタ - カタ)	着 - 角 (チャク - カク)	パイ - 鯛 (パイ - タイ)	貝 - パイ (カイ - パイ)
議事 - 記事 (ギジ - キジ)	ミス - ビス (ミス - ビス)	私利 - 地理 (シリ - チリ)	式 - 引き (シキ - ヒキ)	見栄 - 煮え (ミエ - ニエ)	銀 - 瓶 (ギン - ビン)
銀 - 金 (ギン - キン)	見る - ビル (ミス - ビス)	昼 - 着る (ヒル - キル)	知事 - 記事 (チジ - キジ)	ミス - ニス (ミス - ニス)	気障 - ビザ (キザ - ビザ)
髓 - 粹 (ズイ - スイ)	無理 - 鰯 (ムリ - ブリ)	好き - 月 (スキ - ツキ)	中 - 空 (チュウ - クウ)	剥く - 抜く (ムク - スク)	黒 - プロ (クロ - プロ)
ぐう - 食う (グー - クー)	無視 - 武士 (ムシ - ブシ)	砂 - 綱 (スナ - ツナ)	純 - 群 (ジュン - グン)	無視 - 主 (ムシ - ヌシ)	タ - ルー (ユー - ルー)
税 - 生 (ゼイ - セイ)	面 - 弁 (メン - ベン)	変 - 剣 (ヘン - ケン)	シェア - ヘア (シェア - ヘア)	面 - 年 (メン - ネン)	弦 - 弁 (ゲン - ベン)
出歯 - 手羽 (デバ - テバ)	練る - 出る (ネル - デル)	縁 - けり (ヘリ - ケリ)	シェル - 経る (シェル - ヘル)	ペン - 天 (ペン - テン)	剣 - ペン (ケン - ペン)
像 - 僧 (ゾウ - ソウ)	門 - 盆 (モン - ボン)	星 - 腰 (ホシ - コシ)	条 - 号 (ジョウ - ゴウ)	毛 - 脳 (モウ - ノウ)	語気 - 簿記 (ゴキ - ボキ)
誤字 - 孤児 (ゴジ - コジ)	野良 - 銅鑼 (ノラ - ドラ)	掘る - 凝る (ホル - コル)	所持 - 保持 (ショジ - ホジ)	ポロ - とろ (ポロ - トロ)	余暇 - ろ過 (ヨカ - ロカ)

A.2. The answer sheet of listening test.

【実験目的】

音源を聞いて、単語が正しく聞えるかを調べる。

【内容】

以下の種類別で10単語が連続で再生されます。

1単語再生された後、聞き取れた単語を以下の表に書き出してください。

※以下の図のように単語が再生されますので、書き取り時間に注意して回答してください

解答表

	①	②	③	④	⑤	⑥	⑦	⑧	⑨	⑩
Voicing										
Nasality										
Sustention										
Sibilant										
Creveness										
Compactness										