Audio Object Editing Scheme in Spatial Audio Object Coding for User Interaction

Kwangki Kim

1 Korea Nazarene University, Department of Digital Contents, Wolbong-ro 48, Cheonan Chungnam, Korea
k2kim@kornu.ac.kr

Abstract. In this paper, we proposed the object control module in the SAOC has the functionalities such as the object deletion, the insertion and the replacement. For the implementation of the object control module, the down-mix signal and the OLD are simply modified. Using the proposed object control module, a user who enjoys the interactive audio service can eliminate the specific audio object and it can be replaced by another audio object played by the user. In addition, the user can insert the new audio object into a package of the various audio objects. Consequently, the original audio objects content can be modified and the new audio objects content can be distributed.

Keywords: spatial audio object coding, object editing, object control, spatial cue, down-mix signal.

1 Introduction

Recently, an interactive audio service has been introduced to provide users with an enhanced and alternative audio service [1], [2]. The interactive audio service delivers various audio objects such as bass, guitar, and vocal, and so on to users. Therefore, users can control the audio objects and create their own audio sound according to their preferences. The interactive audio service, however, has a constraint that a high bit-rate should be guaranteed because the various audio objects need to be separately coded. To solve the bit-rate problem in the interactive audio service, a spatial audio object coding (SAOC) scheme has been introduced and developed [3], [4].

The SAOC is a technology to represent various audio objects such as vocal, piano, guitar, drum and so on as a down-mix signal and spatial cues. As the SAOC only needs the bit-rate of the down-mix signal and the additional side information, the bit-rate of the interactive audio service can be greatly reduced. The interactive audio service supported by the SAOC can adopt a new application scenario such as the deletion and the replacement of the existing specific audio object or the addition of a new audio object. In this application scenario, any user can eliminate the specific audio object from the provided whole audio objects and it can be replaced by another audio object played by the user. In addition, the user can insert the new audio object into a package of the various audio objects. From this process, the original audio objects content can be modified and the new audio objects content can be distributed.
In this paper, we proposed an object control module for the above application scenario in the interactive audio service supported by the SAOC.

2 Overview of the SAOC

In the SAOC encoder, the input audio objects are down-mixed and the spatial cues are extracted to represent the object information. In the SAOC decoder, the audio objects are reconstructed by the transmitted down-mix signal and the spatial cues, reversely. Then, the recovered audio objects are inputted to a renderer and the desired output audio signal is generated according to the user’s preference.

The SAOC uses an object level difference (OLD) parameter as the spatial cue. The OLD indicates power information of each audio object. The OLD can be calculated as

\[
OLD_{(pb)} = \frac{P_i(p_b)}{P_{\text{max}}(p_b)} \quad 1 \leq i \leq N, 1 \leq p_b \leq M
\]

where \(P_i(p_b)\) and \(P_{\text{max}}(p_b)\) are the estimated power of the \(i^{th}\) audio object and the maximum power among the input audio objects at the sub-band \(p_b\), respectively [5]. \(N\) and \(M\) are the numbers of the input audio objects and the sub-bands, respectively.

3 Proposed Object Control Module

Fig. 1 shows an overall structure of the SAOC with the object control module. The object control module takes the down-mix signal, the SAOC bit-stream, control information, and the inserted or changed object as inputs and it generates the modified down-mix signal and the SAOC bit-stream through the object deletion, insertion and the replacement processes according to the control information.

3.1 Object deletion

The object deletion consists of two processes such as the down-mix generation and the parameter estimation as shown in Fig. 2. In the down-mix generation, when the \(i^{th}\)
specific audio object is eliminated, the modified down-mix signal is generated by subtracting the power of the $i$\textsuperscript{th} audio object from the power of the down-mix signal using (2).

$$\hat{X}_j(k) = X_j(k) \left[1 - \text{OLD}_j \sum_{j=1}^{\text{OLD}_j} \right]$$

(2)

where $\hat{X}_j(k)$ and $X_j(k)$ are the modified and the original down-mix signal in the frequency domain, respectively. $j$ and $k$ are the object number index and the frequency bin, respectively. In the parameter estimation, the OLD parameters are re-estimated according to value of $\text{OLD}_j$ to be the OLD of the eliminated object. If $\text{OLD}_j$ is not 1, the OLD of the $i$\textsuperscript{th} audio object is just eliminated from the SAOC bit-stream and the OLDS of the other audio object are not changed. So, there is no a further process for the OLD estimation. Otherwise, the OLD parameters are re-calculated by dividing all OLDS into the maximum OLD among the remained OLDS as (3).

$$\tilde{\text{OLD}}_n(pb) = \frac{\text{OLD}_n(pb)}{\text{OLD}_{\text{max}}(pb)} \left[1 \leq n \leq N - 1 \right] \left[1 \leq pb \leq M \right]$$

(3)

where $\tilde{\text{OLD}}_n(pb)$ and $\text{OLD}_{\text{max}}(pb)$ are the nth re-estimated OLD and the maximum OLD among the remained OLDS at the sub-band $pb$, respectively.

### 3.2 Object insertion

The object deletion also consists of two processes such as the down-mix generation and the parameter estimation as shown in Fig. 3. In the down-mix generation, the inserted audio object is simply added into the down-mix signal as (4).

$$\hat{X}_j(k) = X_j(k) + X_{\text{new}}(k)$$

(4)

where $X_{\text{new}}(k)$ is the inserted audio object in the frequency domain. For the parameter estimation, the powers of all audio objects are firstly calculated using the original OLDS and the down-mix signal as (5).
\[ \hat{P}(pb) = P_i(pb) \sum_{i=1}^{N} \frac{OLD_i(pb)}{OLD_j(pb)} \quad \text{for} \ 1 \leq i \leq N \]  

where \( \hat{P}(pb) \) and \( P_i(pb) \) are the estimated power of the \( i \)th audio object and the power of the down-mix signal at the sub-band \( pb \), respectively. Then, the new OLDs are re-estimated with the inserted audio object and the estimated audio objects using (1). Here, since the number of the audio objects increases, the number of the estimated OLDs is \( N+1 \).

3.3 Object replacement

The object replacement consists of the combination of the object deletion and the object insertion. For the modification of the down-mix signal, the specific audio object is extracted from the down-mix signal using (2) and the new audio object is added into the modified down-mix signal using (4). For the modification of the OLD, the powers of all audio objects are calculated with the down-mix signal and the OLDs using (5). Then, the new OLDs are estimated with the estimated powers of all audio objects except the deleted audio object and the power of the new inserted audio object using (1).

4 Conclusion

In this paper, we proposed the object control module in the SAOC has the functionalities such as the object deletion, the insertion and the replacement. For the implementation of the object control module, the down-mix signal and the OLD are simply modified. Using the proposed object control module, the user who enjoys the interactive audio service can eliminate the specific audio object and it can be replaced by another audio object played by the user. In addition, the user can insert the new audio object into a package of the various audio objects. Consequently, the original audio objects content can be modified and the new audio objects content can be distributed. As the future work, the performance of the proposed object control module should be validated, firstly. And the expected sound degradation problem by the object deletion will be studied.

References
