codec
Internet-Draft
Updates: 7845 (if approved)
Intended status: Standards Track
Expires: December 21, 2018

J. Skoglund
Google LLC
M. Graczyk

June 19, 2018

Ambisonics in an Ogg Opus Container
draft-ietf-codec-ambisonics-07

Abstract

This document defines an extension to the Opus audio codec to encapsulate coded ambisonics using the Ogg format. It also contains updates to RFC 7845 to reflect necessary changes in the description of channel mapping families.

Status of This Memo

Copyright Notice

Table of Contents

1. Introduction

Ambisonics is a representation format for three dimensional sound fields which can be used for surround sound and immersive virtual reality playback. See [gerzon75] and [daniel04] for technical details on the ambisonics format. For the purposes of the this document, ambisonics can be considered a multichannel audio stream. A separate stereo stream can be used alongside the ambisonics in a head-tracked virtual reality experience to provide so-called non-diegetic audio - audio which should remain unchanged by listener head rotation; e.g., narration or stereo music. Ogg is a general purpose container, supporting audio, video, and other media. It can be used to encapsulate audio streams coded using the Opus codec. See [RFC6716] and [RFC7845] for technical details on the Opus codec and its encapsulation in the Ogg container respectively.

This document extends the Ogg Opus format by defining two new channel mapping families for encoding ambisonics. The Ogg Opus format is extended indirectly by adding an item with value 2 or 3 to the IANA "Opus Channel Mapping Families" registry. When 2 or 3 are used as the Channel Mapping Family Number in an Ogg stream, the semantic meaning of the channels in the multichannel Opus stream is one of the ambisonics layouts defined in this document. This mapping can also be used in other contexts which make use of the channel mappings defined by the Opus Channel Mapping Families registry. Furthermore, mapping families 240 through 254 (inclusively) are reserved for experimental use.

2. Terminology

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in BCP 14 [RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here.

3. Ambisonics With Ogg Opus

Ambisonics can be encapsulated in the Ogg format by encoding with the Opus codec and setting the channel mapping family value to 2 or 3 in the Ogg identification header (ID). A demuxer implementation encountering Channel Mapping Family 2 or Family 3 MUST interpret the Opus stream as containing ambisonics with the format described in Section 3.1 or Section 3.2, respectively.

3.1. Channel Mapping Family 2

Allowed numbers of channels: (1 + n)^2 + 2j for n = 0...14 and j = 0 or 1, where n denotes the (highest) ambisonic order and j denotes whether or not there is a separate non-diegetic stereo stream. This corresponds to periphonic ambisonics from zeroth to fourteenth order plus potentially two channels of non-diegetic stereo. Explicitly the allowed number of channels are 1, 3, 4, 6, 9, 11, 16, 18, 25, 27, 36, 38, 49, 51, 64, 66, 81, 83, 100, 102, 121, 123, 144, 146, 169, 171, 196, 198, 225, and 227.

This channel mapping uses the same channel mapping table format used by channel mapping family 1. The output channels are ambisonic components ordered in Ambisonic Channel Number (ACN) order, defined in Figure 1, followed by two optional channels of non-diegetic stereo indexed (left, right). The terms order and degree are defined according to [ambix].

$$ACN = n * (n + 1) + m,$$

for order n and degree m.

Figure 1: Ambisonic Channel Number (ACN)

For the ambisonic channels the ACN component corresponds to channel index as k = ACN. The reverse correspondence can also be computed for an ambisonic channel with index k.

```
order   n = floor(sqrt(k)),
degree  m = k - n * (n + 1).
```

Figure 2: Ambisonic Degree and Order from ACN

Note that channel mapping family 2 allows for so-called mixed order ambisonic representation where only a subset of the full ambisonic order number of channels is encoded. By specifying the full number in the channel count field, the inactive ACNs can then be indicated in the channel mapping field using the index 255.

Ambisonic channels are normalized with Schmidt Semi-Normalization (SN3D). The interpretation of the ambisonics signal as well as detailed definitions of ACN channel ordering and SN3D normalization are described in [ambix] Section 2.1.

3.2. Channel Mapping Family 3

Allowed numbers of channels: (1 + n)^2 + 2j for n = 0...14 and j = 0 or 1, where n denotes the (highest) ambisonic order and j denotes whether or not there is a separate non-diegetic stereo stream. This corresponds to periphonic ambisonics from zeroth to fourteenth order plus potentially two channels of non-diegetic stereo. Explicitly the allowed number of channels are 1, 3, 4, 6, 9, 11, 16, 18, 25, 27, 36, 38, 49, 51, 64, 66, 81, 83, 100, 102, 121, 123, 144, 146, 169, 171, 196, 198, 225, and 227.

In this mapping, C output channels (the channel count) are generated at the decoder by multiplying K = N + M decoded channels with a designated demixing matrix, D, having C rows and K columns. Here, N denotes the number of streams encoded and M the number of these which are coupled to produce two channels. As for channel mapping family 2 this mapping family also allows for encoding and decoding of full order ambisonics, mixed order ambisonics, and for non-diegetic stereo channels, but also has the added flexibility of mixing channels. Let X denote a column vector containing K decoded channels X1, X2, ..., XK (from N streams), and let S denote a column vector containing C output streams S1, S2, ..., SC. Then S = D X, i.e.,

```
 /     \   /                   \ /     \
 | S1  |   | D11  D12  ... D1K | | X1  |
 | S2  |   | D21  D22  ... D2K | | X2  |
 | ... | = | ...  ...  ... ... | | ... |
 | SC  |   | DC1  DC2  ... DCK | | XK  |
 \     /   \                   / \     /
```

Figure 3: Demixing in Channel Mapping Family 3

The matrix MUST be provided as side information and MUST be stored in the channel mapping table part of the identification header, c.f. section 5.1.1 in [RFC7845]. The matrix replaces the need for a channel mapping field and for channel mapping family 3 the mapping table has the following layout:


```
  0                   1                   2                   3
  0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
                                                 +-+-+-+-+-+-+-+-+
                                                 | Stream Count  |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 | Coupled Count | Demixing Matrix                               :
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 4: Channel Mapping Table for Channel Mapping Family 3

The fields in the channel mapping table have the following meaning:

1. Stream Count 'N' (8 bits, unsigned):

   This is the total number of streams encoded in each Ogg packet.

2. Coupled Stream Count 'M' (8 bits, unsigned):

   This is the number of the N streams whose decoders are to be configured to produce two channels (stereo).

3. Demixing Matrix (16*K*C bits, signed):

The coefficients of the demixing matrix stored column-wise as 16-bit, signed, two's complement fixed-point values with 15 fractional bits (Q15), little endian. If needed, the output gain field can be used for a normalization scale. For mixed order ambisonic representations, the silent ACN channels are indicated by all zeros in the corresponding rows of the mixing matrix. This allows also for mixed order with non-diegetic stereo as the number of columns implies the presence of non-diegetic channels.

Note that [RFC7845] specifies that the identification header cannot exceed one "page", which is 65,025 octets. This limits the ambisonic order to be lower than 12, if full order is utilized and the number of coded streams is the same as the ambisonic order plus the two non-diegetic channels. The total output channel number, C, MUST be set in the 3rd field of the identification header.

4. Downmixing

The downmixing matrices in this section are only examples known to give acceptable results for stereo downmixing from ambisonics, but other mixing strategies will be allowed, e.g., to emphasize a certain panning.

An Ogg Opus player MAY use the matrix in Figure 5 to implement downmixing from multichannel files using Channel Mapping Family 2 and 3, when there is no non-diegetic stereo. The first and second ambisonic channels are known as "W" and "Y" respectively.

```
 /   \   /                  \ /     \
 | L |   | 0.5  0.5 0.0 ... | |  W  |
 | R | = | 0.5 -0.5 0.0 ... | |  Y  |
```

```
                 \   /   \                  / | ... |
                                              \     /
```

Figure 5: Stereo Downmixing Matrix for Channel Mapping Family 2 and 3 - only Ambisonic Channels

The first ambisonic channel (W) is a mono audio stream which represents the average audio signal over all directions. Since W is not directional, Ogg Opus players MAY use W directly for mono playback.

If a non-diegetic stereo track is present, the player MAY use the matrix in Figure 6 for downmixing. Ls and Rs denote the two non-diegetic stereo channels.

```
 /   \   /                              \  /     \
 | L |   | 0.25  0.25 0.0 ... 0.5 0.0 |  |  W   |
 | R | = | 0.25 -0.25 0.0 ... 0.0 0.5 |  |  Y   |
 \   /   \                              /  | ... |
                                           |  Ls |
                                           |  Rs |
                                           \     /
```

Figure 6: Stereo Downmixing Matrix for Channel Mapping Family 2 and 3 - Ambisonic Channels Plus a Non-diegetic Stereo Stream

5. Updates to RFC 7845

5.1. Format of the Channel Mapping Table

The language in section 5.1.1 in [RFC7845] implies that the channel mapping table, when present, has a fixed format for all channel mapping families:

> The order and meaning of these channels are defined by a channel mapping, which consists of the 'channel mapping family' octet and, for channel mapping families other than family 0, a 'channel mapping table', as illustrated in Figure 3.

This document updates [RFC7845] to clarify that the format of the channel mapping table may depend on the channel mapping family:

The order and meaning of these channels are defined by a channel mapping, which consists of the 'channel mapping family' octet and for channel mapping families other than family 0, a 'channel mapping table'.

The format of the channel mapping table depends on the channel mapping family. Unless the channel mapping family requires a custom format for its channel mapping table, the RECOMMENDED channel mapping table format for new mapping families is illustrated in Figure 3.

The change above is not meant to change how families 1 and 255 currently work. To ensure that, the first paragraph of Section 5.1.1.2 is changed from:

Allowed numbers of channels: 1...8. Vorbis channel order (see below).

to

Allowed numbers of channels: 1...8, with the mapping specified according to Figure 3. Vorbis channel order (see below).

Similary, the first paragraph of Section 5.1.1.4 is changed from:

Allowed numbers of channels: 1...255. No defined channel meaning.

to

Allowed numbers of channels: 1...255, with the mapping specified according to Figure 3. No defined channel meaning.

5.2. Unknown Mapping Families

The treatment of unknown mapping families is changed slightly. Section 5.1.1.4 of [RFC7845] states:

The remaining channel mapping families (2...254) are reserved. A demuxer implementation encountering a reserved 'channel mapping family' value SHOULD act as though the value is 255.

This is changed to:

The remaining channel mapping families (2...254) are reserved. A demuxer implementation encountering a 'channel mapping family' value that it does not recognize SHOULD NOT attempt to decode the packets and SHOULD NOT use any information except for the first 19 octets of the ID header packet (Fig. 2) and the comment header (Fig. 10).

6. Experimental Mapping Families

To make development of new mapping families easier while reducing the risk of creating compatibility issues with non-final version of mapping families, mapping families 240 through 254 (inclusively) are now reserved for experiments and implementations of in-development families. Implementers SHOULD attempt to use experimental family numbers that have not recently been used and SHOULD advertise what experimental numbers they use (e.g. for Internet-Drafts).

The ambisonics mapping experiments that led to this document used experimental family 254 for family 2 and experimental family 253 for family 3.

7. Security Considerations

Implementations of the Ogg container need to take appropriate security considerations into account, as outlined in Section 10 of [RFC7845]. The extension defined in this document requires that semantic meaning be assigned to more channels than the existing Ogg format requires. Since more allocations will be required to encode and decode these semantically meaningful channels, care should be taken in any new allocation paths. Implementations MUST NOT overrun their allocated memory nor read from uninitialized memory when managing the ambisonic channel mapping.

8. IANA Considerations

This document updates the IANA Media Types registry "Opus Channel Mapping Families" to add 17 new assignments.

Value	Reference
2	This Document Section 3.1
3	This Document Section 3.2
240-254	This Document Section 6

9. Acknowledgments

Thanks to Timothy Terriberry, Jean-Marc Valin, Mark Harris, Marcin Gorzel, and Andrew Allen for their guidance and valuable contributions to this document.

10. References

10.1. Normative References

[ambix] Nachbar, C., Zotter, F., Deleflie, E., and A. Sontacchi, "AMBIX - A SUGGESTED AMBISONICS FORMAT", June 2011, <http://iem.kug.ac.at/fileadmin/media/iem/projects/2011/ambisonics11_nachbar_zotter_sontacchi_deleflie.pdf>.

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <http://www.rfc-editor.org/info/rfc2119>.

[RFC6716] Valin, JM., Vos, K., and T. Terriberry, "Definition of the Opus Audio Codec", RFC 6716, DOI 10.17487/RFC6716, September 2012, <http://www.rfc-editor.org/info/rfc6716>.

[RFC7845] Terriberry, T., Lee, R., and R. Giles, "Ogg Encapsulation for the Opus Audio Codec", RFC 7845, DOI 10.17487/RFC7845, April 2016, <http://www.rfc-editor.org/info/rfc7845>.

[RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174, May 2017, <https://www.rfc-editor.org/info/rfc8174>.

10.2. Informative References

[daniel04]
Daniel, J. and S. Moreau, "Further Study of Sound Field Coding with Higher Order Ambisonics", May 2004, <http://pcfarina.eng.unipr.it/Public/phd-thesis/aes116%20high-passed%20hoa.pdf>.

[gerzon75]
Gerzon, M., "Ambisonics. Part one: General system description", August 1975, <http://www.michaelgerzonphotos.org.uk/articles/Ambisonics%201.pdf>.

Authors' Addresses

Jan Skoglund
Google LLC
345 Spear Street
San Francisco, CA 94105
USA

Email: jks@google.com

Michael Graczyk

Email: michael@mgraczyk.com