codec
Internet-Draft
Intended status: Standards Track
Expires: September 28, 2017

J. Skoglund
M. Graczyk
Google Inc.
March 27, 2017

Ambisonics in an Ogg Opus Container
draft-ietf-codec-ambisonics-02

Abstract

This document defines an extension to the Ogg format to encapsulate ambisonics coded using the Opus audio codec.

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at http://datatracker.ietf.org/drafts/current/.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on September 28, 2017.

Copyright Notice

Copyright (c) 2017 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (http://trustee.ietf.org/license-info) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1. Introduction 2
2. Terminology 2
3. Ambisonics With Ogg Opus 3
3.1. Channel Mapping Family 2 3
3.2. Channel Mapping Family 3 4
4. Downmixing 5
5. Security Considerations 6
6. IANA Considerations 6
7. Acknowledgments 7
8. References 7
8.1. Normative References 7
8.2. Informative References 7
Authors' Addresses 8

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 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 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.

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 [RFC2119].

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 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, 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).

```
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 representations where only a subset of the full ambisonic order number of channels is used. 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 expected to be 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...12 and j = 0 or 1, where n denotes the (highest) ambisonic order and j whether or not there is a separate non-diegetic stereo stream. This corresponds to periphonic ambisonics from zeroth to twelfth 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.

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). 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 demixing matrix.

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. Also note that the total output channel number, C, MUST be set in the 3rd field of the identification header.

4. Downmixing

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. This downmixing is known to give acceptable results for stereo downmixing from ambisonics. 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. Security Considerations

Implementations of the Ogg container need 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.

6. IANA Considerations

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

Value	Reference
2	This Document Section 3.1
3	This Document Section 3.2

7. Acknowledgments

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

8. References

8.1. Normative References

[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>.

[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>.

8.2. Informative References

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

[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>.

Authors' Addresses

Jan Skoglund
Google Inc.
1600 Amphitheatre Parkway
Mountain View, CA 94043
USA

Email: jks@google.com

Michael Graczyk
Google Inc.
1600 Amphitheatre Parkway
Mountain View, CA 94043
USA

Email: mgraczyk@google.com