

BLE RDK Voice Service Specification

RDK-SP-BLE-RV-Service-D02-200120

Document Status: Draft

January 20, 2020

Document Status

Document Control Number:	RDK-SP-BLE-RV-Service-D02-200120
Document Title:	BLE RDK Voice Service Specification
Versions:	D01 – June 20, 2019 D02 – January 20, 2019
Date:	January 20, 2020
Status:	Document Status: Draft
Distribution:	RDK members only

Document Status Codes

Work in Progress (W) An incomplete document designed to guide discussion and

generate feedback that may include several alternative

requirements for consideration.

Draft (D) A document in specification format considered largely

complete, but lacking review. Drafts are susceptible to

substantial change during the review process.

Issued (I) A stable document that has undergone rigorous review

and is suitable for product design and development. It will

serve as a basis for testing requirements.

January 20, 2020 Page 2 of 24

Table of Contents

1.	Intro	duction	6
	1.1	Overview	6
	1.2	Purpose of Document	6
	1.3	Typographical Conventions	6
	1.4	Revision History	6
2	Refer	ences	7
	2.1	Normative References	7
3	Term	s and Definitions	8
4	Abbr	eviations and Acronyms	9
5	Intro	duction	. 10
	5.1	Conformance	. 10
	5.2	Service Dependency	. 10
	5.3	Bluetooth Specification Release Compatibility	. 10
	5.4	GATT Sub-Procedure Requirements	. 10
	5.5	Transport Dependencies	. 10
	5.6	Error Codes	. 10
	5.7	Byte Transmission Order	. 11
6	Servi	ce Requirements	. 12
	6.1	Service Declaration	. 12
	6.2	Service Roles	. 12
	6.3	Service Sequence Examples	. 12
	6.3.1	Push-to-talk Voice Session Sequence	. 12
	6.3.2	Explicitly Rejected Push-to-talk Voice Session Sequence	. 14
	6.3.3	Far-Field Voice Session Sequence	. 14
	6.4	Characteristic Overview	. 16
	6.5	Audio Codecs Characteristic	. 16
	6.5.1	Audio Codecs Characteristic Value	. 17
	6.6	Audio Gain Characteristic	. 17
	6.6.1	Audio Gain Characteristic Value	. 17

6.7	Αι	ıdio Control Characteristic	17
(6.7.1	Audio Control Characteristic Value	18
(6.7.2	Audio Control Encoding Value	18
(6.7.3	Audio Control Enable Value	18
(6.7.4	Audio Control Enable Behaviour	19
6.8	Αι	ıdio Response Characteristic	19
(6.8.1	Audio Response Characteristic Value	19
(6.8.2	Audio Response Characteristic Behaviour	19
6.9	Ke	yword Detect Characteristic	19
(6.9.1	Keyword Detect Characteristic Descriptors	20
	6.9.1	.1 Client Characteristic Configuration Descriptor	20
(6.9.2	Keyword Detect Characteristic Value	20
(6.9.3	Keyword Detect Characteristic Behaviour	21
6.1	0 Ве	eamformer Data Characteristic	21
(6.10.1	Beamformer Data Characteristic Value	21
	6.10.	1.1 Beam Data Field	21
	6.10.	1.2 Client Characteristic Configuration Descriptor	22
(6.10.2	Beamformer Data Characteristic Behaviour	22
6.1	1 Au	ıdio Data Characteristic	22
(6.11.1	Audio Data Characteristic Descriptors	22
	6.11.	1.1 Client Characteristic Configuration Descriptor	22
(6.11.2	Audio Data Characteristic Value	22
	6.11.	2.1 G.726 and IMA/DVI ADPCM Frame Format	22
	6.11.	2.2 Audio Frame Buffering	24
Table	S		
Table :	1 - Typo	graphical Conventions	6
Table 2	2 - Term	s and Definitions	8
Table 3	3 - Abbr	eviations and Acronyms	9
Table 4	4 - GAT1	Sub-Procedure Requirement	10
Table !	5 - Voice	Service Characteristics	16

Table 6 - Audio Codecs Supported Bit Mask	17
Table 7 - Audio Control Characteristic Value	18
Table 8 - Audio Control Characteristic Encoding Values	18
Table 9 - Audio Control Characteristic Enable Values	18
Table 10 - Audio Response Values	19
Table 11 - Audio Keyword Detect Characteristic	20
Table 12 - Audio Beamformer Characteristic	21
Table 13 - Beam Data Field	21
Table 14 - Beam Description Byte	21
Table 15 - Audio Frame Metadata Format	23
Figures	
Figure 2 – Push-to-talk Voice Session	13
Figure 3 - Explicit Reject of Voice Session	14
Figure 4 - Far-Field Voice Session	16
Figure 5 - G.726 and IMA/DVI ADPCM Frame Format	23
Figure 6 - G.726 and IMA/DVI ADPCM Frame Metadata	23

1. Introduction

1.1 Overview

The RDK Voice Service is a service for transmitting voice data over BLE in the RDK ecosystem. The RDK Voice Service is specifically designed to enable voice control of a BLE voice client.

1.2 Purpose of Document

This document defines detailed requirements for the RDK Voice Service. It is intended to specify transmission of voice data from an RVS server to an RVS client.

1.3 Typographical Conventions

This specification uses different typefaces to differentiate and emphasize important information.

Typeface Usage **Boldface** Used to call attention to a piece of information. For example: This specification does **not** include headend diagnostic screens. **Boldface &** Used to emphasize information and for readability. For example: Uppercase ENTER, MUTE, INFO, VOL +/- and other buttons on the remote control. Italics Used to emphasize that the information being presented is for informational purposes only and is not a requirement even though it may contain conformance language. For example: Note: The voice controller uses the Channel Check Request to verify that the voice target has disabled frequency agility. Used to define and signify a requirement. For example: Uppercase MUST, SHOULD, and MAY.

Table 1 - Typographical Conventions

1.4 Revision History

Version	Date	Author	Remarks
11)()1	09 Aug 2019	Comcast	Initial Version
1)()2	20 Jan 2020	Comcast	Clean-up

January 20, 2020 Page 6 of 24

2 References

Reasonable effort is made to keep references up to date with respect to versions and release dates, however manufacturers are responsible for ensuring they have the most recent version of a reference specification (unless otherwise noted).

Where conflicts exist between requirements contained in this specification and normative references, the specification requirements govern.

2.1 Normative References

[ADPCM] Recommended Practices for Enhancing Digital Audio Compatibility in

Multimedia Systems, IMA Digital Audio Focus and Technical Working

Groups; DATWG Recommendation, October 21, 1992.

[BLUETOOTH] Bluetooth Core Specification version 4.0 or later

January 20, 2020 Page 7 of 24

3 Terms and Definitions

This document uses the following terms and definitions.

Table 2 - Terms and Definitions

Term	Definition
Adaptive Differential Pulse- Code Modulation	A compression algorithm that varies the size of the quantization step, allowing further reduction of the required bandwidth for a given signal-to-noise ratio.
Opus	

January 20, 2020 Page 8 of 24

4 Abbreviations and Acronyms

This document uses the following abbreviations and acronyms.

Table 3 - Abbreviations and Acronyms

Abbrv	Acronym
ADPCM	Adaptive Differential Pulse-Code Modulation
RVS	RDK Voice Service

January 20, 2020 Page 9 of 24

5 Introduction

The RDK Voice Service exposes data and associated formatting for streaming voice audio from an RDK Remote Control Device to an RDK based STB.

5.1 Conformance

All capabilities indicated as mandatory for this Service shall be supported in the specified manner (process-mandatory). This also applies for all optional and conditional capabilities for which support is indicated.

5.2 Service Dependency

This service is not dependent upon any other services.

5.3 Bluetooth Specification Release Compatibility

This specification is compatible with any Bluetooth core specification as defined in [BLUETOOTH] that includes the Generic Attribute Profile (GATT) specification and the Bluetooth Low Energy Controller specification.

5.4 GATT Sub-Procedure Requirements

Requirements in this section represent a minimum set of requirements for an RDK Remote Control Device (GATT Server). Other GATT sub-procedures may be used if supported by both Client and Server.

Table 4 below summarises additional GATT sub-procedure requirements beyond those required by all GATT Servers.

GATT Sub-Procedure	Requirement
Read Characteristic Value	M
Write Characteristic Value	M
Write Without Response	0
Notification	M
Read Characteristic Descriptors	M
Write Characteristic Descriptors	M

Table 4 - GATT Sub-Procedure Requirement

5.5 Transport Dependencies

The service shall only operate over an LE transport.

5.6 Error Codes

This service does not define any application error codes that are used in Attribute Protocol.

January 20, 2020 Page 10 of 24

5.7 Byte Transmission Order

All characteristics used with this service shall be transmitted with the least significant octet first (i.e., little endian).

January 20, 2020 Page 11 of 24

6 Service Requirements

6.1 Service Declaration

The service UUID shall be set to:

0000F800-BDF0-407C-AAFF-D09967F31ACD

6.2 Service Roles

A remote control or similar low power device enabled with one or multiple microphones should function as an RVS Server.

A settop box or other host capable of processing transmitted voice data should function as an RVS Client.

6.3 Service Sequence Examples

6.3.1 Push-to-talk Voice Session Sequence

Figure 1 illustrates a sequence of a normal push-to-talk voice session.

The RVS Server should depend on the application defined HID keypress to enable audio streaming from an RVS Client.

January 20, 2020 Page 12 of 24

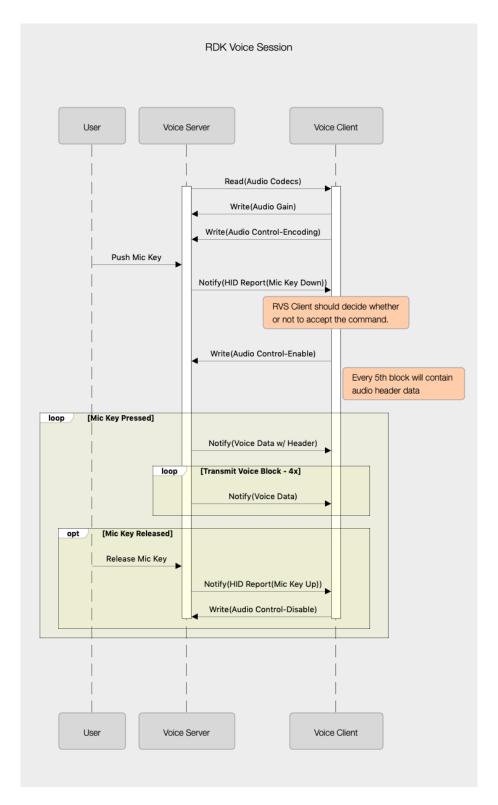


Figure 1 – Push-to-talk Voice Session

January 20, 2020 Page 13 of 24

6.3.2 Explicitly Rejected Push-to-talk Voice Session Sequence

The RVS provides an optional characteristic for the RVS Client to explicitly inform the RVS Server a voice request was rejected and why. Figure 2 illustrates this sequence.

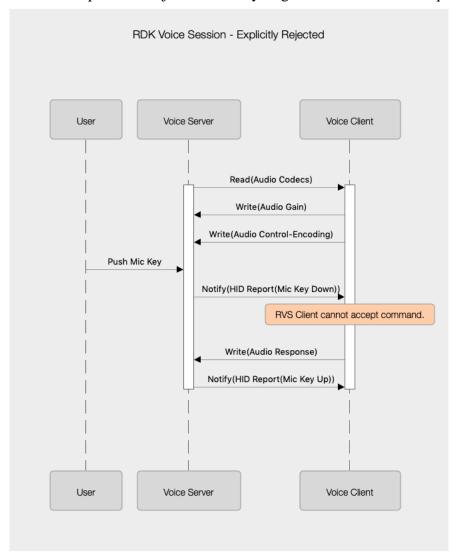


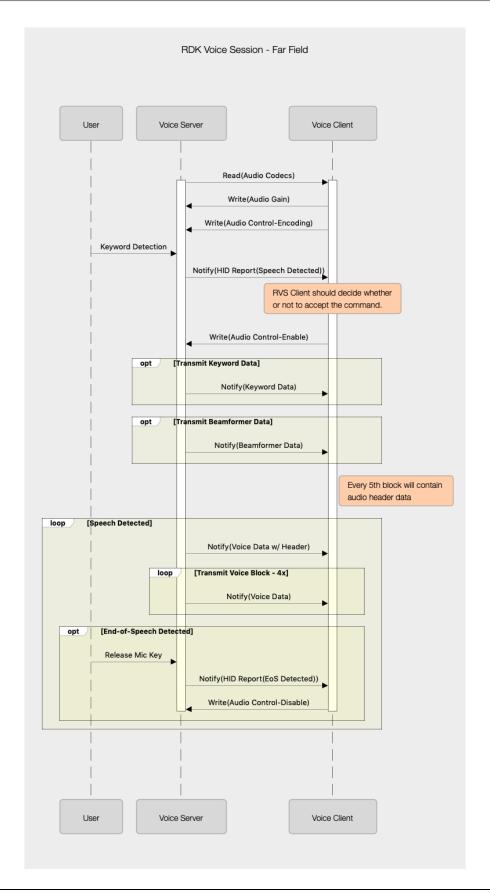
Figure 2 - Explicit Reject of Voice Session

6.3.3 Far-Field Voice Session Sequence

The RVS supports optional characteristics to support transferring keyword and beamformer data to an RVS Client for more optimal processing of the audio stream and metrics.

Figure 3 shows a voice session for a far-field device supporting a keyword detector and end-of-speech detection.

January 20, 2020 Page 14 of 24



January 20, 2020 Page 15 of 24

Figure 3 - Far-Field Voice Session

6.4 Characteristic Overview

The RDK Voice Service is composed of the following characteristics used to provide access to the audio capabilities and data.

Only one instance of each characteristic shall be permitted within an RDK Voice Service unless otherwise specified.

Characteristic Requirement Mandatory Optional Security Permissions Name **Properties Properties Audio Codecs** M Read None 0 Audio Gain Read. Write. None Write Without Response **Audio Control** Μ Read, Write, None Write Without Response Audio Response 0 Write None **Keyword Data** 0 Notify None Beamformer Data 0 Notify None Audio Data M Notify None

Table 5 - Voice Service Characteristics

Notes:

- Security Permissions of "None" means that this service does not impose any requirements.
- Profiles utilising this Service may impose security requirements beyond those defined in Table
 2.1 for all characteristics defined in Table 2.1.
- Properties not listed as mandatory (M) or optional (O) are excluded.

Profiles utilising this Service may impose security requirements beyond those defined in Table 5 for all characteristics defined in Table 5.

Properties not listed as mandatory (M) or optional (O) shall be excluded.

6.5 Audio Codecs Characteristic

The Audio Codecs characteristic is used to expose the codecs supported by an RVS Server.

This characteristic shall be read only.

The Audio Codecs characteristic shall be written by the RVS Server.

January 20, 2020 Page 16 of 24

Only a single instance of this characteristic shall exist as part of the RDK Voice Service.

The characteristic UUID shall be set to:

0000EA00-BDF0-407C-AAFF-D09967F31ACD

6.5.1 Audio Codecs Characteristic Value

The Audio Codecs Characteristic value is a 32-bit little endian value that contains a bit mask of supported codecs.

Only the codecs shown in Table 6 shall be set.

An RVS Server shall support at least one codec shown in the table.

Table 6 - Audio Codecs Supported Bit Mask

Bit	Audio Codec	Bits per	Audio Sample	Audio
		Sample	Rate	Channels
0	G.726-32 ADPCM	4	16000 per second	Single (Mono)
1	IMA/DVI ADPCM	4	16000 per second	Single (Mono)
2	Opus			Single (Mono)
2 -	Reserved for future	-	-	-
31	use			

6.6 Audio Gain Characteristic

The Audio Gain characteristic is used to expose the gain level of the microphone used for voice capture.

Only a single instance of this characteristic shall exist as part of the RDK Voice Service.

The characteristic UUID of the Audio Gain shall be set to:

0000EA01-BDF0-407C-AAFF-D09967F31ACD

6.6.1 Audio Gain Characteristic Value

The Audio Gain Characteristic value is an unsigned 8-bit value that contains the current audio gain value. The minimum value is 0 and the maximum value is 64.

The Audio Gain characteristic value shall be persistent across connections for bonded devices. The default value for the Audio Gain Characteristic value is vendor specific. Upon connection of non-bonded clients, this characteristic value is set to the default value.

6.7 Audio Control Characteristic

The Audio Control Characteristic value contains two fields; the first signals the audio codec to be used for the encoding and the second is for enablement / disablement of the audio recording.

The Audio Control Characteristic value can be read using either the GATT Read Characteristic Value and is written using the GATT Write Characteristic Value or optional GATT Write Without Response sub-procedure.

January 20, 2020 Page 17 of 24

Only a single instance of this characteristic shall exist as part of the RDK Voice Service.

The characteristic UUID shall be set to:

0000EA02-BDF0-407C-AAFF-D09967F31ACD

6.7.1 Audio Control Characteristic Value

The Audio Control Characteristic value shall be reset to the default value following connection establishment.

Table 7 shows the data format of the characteristic value.

Table 7 - Audio Control Characteristic Value

Name	Requirement	Format	Default Value
Audio Encoding	Mandatory	uint8	0x00
Audio Enable	Mandatory	uint8	0x00

6.7.2 Audio Control Encoding Value

The Audio Encoding setting of the Audio Control Characteristic shall only take effect the next time audio recording is started.

Changing the Audio Encoding value shall be ignored by the remote device if audio streaming is currently enabled, however reading this value shall always contain the last value written.

Table 8 shows the possible values for the Audio Encoding value.

Table 8 - Audio Control Characteristic Encoding Values

Value	Description
0	Audio is to be encoded using the G.726 ADPCM codec
1	Audio is to be encoded using the IMA/DVI ADPCM codec
2	Audio is to encoded using the Opus codec.
2 - 255	Reserved for future use

The RVS Client shall ensure that only codecs reported in the Audio Control Characteristic as available will be set in the Audio Control Encoding Value.

6.7.3 Audio Control Enable Value

The Audio Enable setting of the Audio Control Characteristic shall have only the values shown in Table 9.

Table 9 - Audio Control Characteristic Enable Values

Value	Description
0	Disable audio streaming

January 20, 2020 Page 18 of 24

1	Enable audio streaming
2 - 255	Reserved for future use

6.7.4 Audio Control Enable Behaviour

When the Audio Enable Setting is toggled to 0×01 then audio data shall start to be sent via the Audio Data Characteristic provided notifications are enabled for the characteristic.

6.8 Audio Response Characteristic

The Audio Response characteristic is used to allow the RVS Client to provide a reason to the RVS Server for not initiating a voice session.

The Audio Response characteristic is optional.

The characteristic UUID shall be set to **.

6.8.1 Audio Response Characteristic Value

The Audio Response can have the values shown in Table 9.

Table 10 - Audio Response Values

Session Response Enum	Value	Description
Busy	0x1	The device is performing other tasks that prevent it from beginning a voice session. An example of this could be another voice session from a different device is in progress already or a captive firmware update is in progress.
Voice Server Not Ready	0x2	An associated endpoint to send voice data to is not available for a voice session.
Not Supported	0x3	Device does not support voice control.
Failure	0x4	An unspecified failure has occurred on the device preventing it from beginning a voice session.
Reserved	0x5-0xFF	Reserved for future use.

6.8.2 Audio Response Characteristic Behaviour

If the RVS Server supports the Audio Response characteristic, the RVS Client MAY write the Audio Response Characteristic if it is unable to accept notifies on the Audio Data characteristic.

6.9 Keyword Detect Characteristic

The Keyword Detect characteristic is used to allow the RVS Client to provide information to the RVS Server about a keyword detect that initiated the voice session.

The Audio Keyword Detect characteristic is optional.

January 20, 2020 Page 19 of 24

The characteristic UUID shall be set to:

TBD

6.9.1 Keyword Detect Characteristic Descriptors

6.9.1.1 Client Characteristic Configuration Descriptor

A Client Characteristic Configuration descriptor shall be included in the Keyword Detect characteristic.

6.9.2 Keyword Detect Characteristic Value

The Keyword Detect Characteristic contains 6 fields specifying information about a keyword detection that initiated the voice session as described in Table 11.

Table 11 - Audio Keyword Detect Characteristic

Field	Size	Description
Pre-Keyword Sample Qty	4 Octets	The Pre-Keyword Sample Qty is a value that represents the number of samples available prior to the start of the keyword.
Keyword Sample Qty	4 Octets	The Keyword Sample Qty is a value that represents the number of samples contained in the keyword.
Estimated Direction of Arrival	2 Octets	The Estimated Direction of Arrival is a value that represents the angle at which the audio is being captured from (0-360 deg).
Standard Search Point	1 Octet	The Standard Search Point represents the sensitivity of the first level keyword detector.
High Search Point	1 Octet	The High Search Point represents the sensitivity of the second level keyword detector. This value can be set to 0xFF if the second level keyword detector does not exist or is disabled.
Dynamic Gain	1 Octet	The Dynamic Gain represents the amount of gain applied to the audio stream.

January 20, 2020 Page 20 of 24

6.9.3 Keyword Detect Characteristic Behaviour

If the RVS Server supports the Audio Keyword Detect characteristic, the RVS Server MAY notify on the Audio Keyword Detect Characteristic if the audio session was initiated by a keyword before the Audio Data characteristic is notified on.

6.10Beamformer Data Characteristic

The Audio Beamformer Data characteristic is used to allow the RVS Server to provide information about beamformers used in the voice session.

The Audio Beamformer Data characteristic is optional.

The characteristic UUID shall be set to:

TBD

6.10.1Beamformer Data Characteristic Value

The Beamformer Data Characteristic contains up to 4 fields describing beam data from a beamformer running on the RVS Client as described in Table 12.

Table 12 - Audio Beamformer Characteristic

Field	Size	Description
Beam Data 1	5 Octets	Beam 1 data.
Beam Data 2	5 Octets	Beam 2 data.
Beam Data 3	5 Octets	Beam 3 data.
Beam Data 4	5 Octets	Beam 4 data.

6.10.1.1 Beam Data Field

The format used for each Beam Data field in the Beamformer Data Characteristic is describes in Table 13.

Table 13 - Beam Data Field

Bits 8	16	16
Beam Description	Confidence	Signal Noise Ratio

The Beam Description byte uses the format described in Table 14.

Table 14 - Beam Description Byte

Bits 0-3	Bit 4	Bit 5	Bits 6-7
Reserved	Selected	Triggered	Angle (0, 90, 270, 360)

January 20, 2020 Page 21 of 24

The Confidence value is float between 0-1 that has been converted to an unsigned 16-bit integer. This value represents the confidence level of the keyword detector.

The Signal Noise Ratio is a signed 16-bit integer where 1 unit is equal to .01 SNR.

6.10.1.2 Client Characteristic Configuration Descriptor

A Client Characteristic Configuration descriptor shall be included in the Beamformer Data characteristic.

6.10.2 Beamformer Data Characteristic Behaviour

If the RVS Server supports the Beamformer Data characteristic, the RVS Server MAY notify on the Beamformer Data Characteristic if the audio session was initiated by a keyword before the Data characteristic is notified on.

6.11Audio Data Characteristic

The Audio Data characteristic is used to send data from an RVS Server to an RVS Client.

Only a single instance of this characteristic shall exist as part of the RDK Voice Service. The characteristic UUID shall be set to:

0000EA03-BDF0-407C-AAFF-D09967F31ACD

6.11.1Audio Data Characteristic Descriptors

6.11.1.1 Client Characteristic Configuration Descriptor

A Client Characteristic Configuration descriptor shall be included in the Audio Data characteristic.

6.11.2Audio Data Characteristic Value

The Audio Data Characteristic value consists of 20 octets of encoded audio data.

When the audio data streaming is enabled via the Audio Control Characteristic and notifications are enabled in the Client Characteristic Configuration descriptor then 20 octets of Audio Data shall be sent as a GATT notification.

Audio data shall be stored in frames, each frame is a multiple of 20 octets and will be segmented and sent in order within a GATT notification. The format and size of the frame depends on the codec used for encoding the audio data. Only complete frames shall be sent in fixed sized batches of GATT notifications. If the Remote Control Device has to discard data (i.e. in noisy environments) it shall discard a complete audio frame.

6.11.2.1 G.726 and IMA/DVI ADPCM Frame Format

Both G.726 and IMA/DVI ADPCM Frames shall be 100 octets in length; consisting of 96 octets of encoded audio sample data and 4 octets of metadata.

January 20, 2020 Page 22 of 24

Figure 4 shows the layout of an audio frame and the how it is sent in 20 octet GATT notifications.

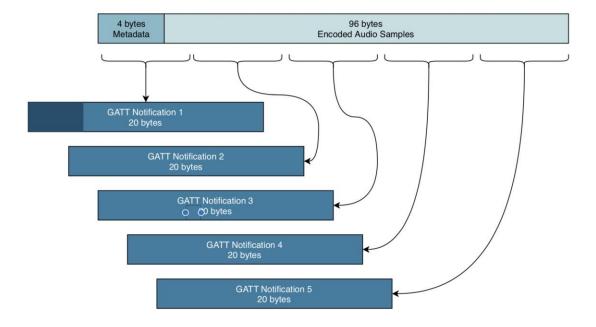


Figure 4 - G.726 and IMA/DVI ADPCM Frame Format

The 4 octet meta data header contains information to be used by the decoder, Table 15 shows the format of the meta data.

Table 15 - Audio Frame Metadata Format

Octet	Description
0	Frame sequence number
1	Index into stepsize table for the start of the frame
2-3	Predicted value of first sample from the previous frame (little endian byte order)

Figure 5 shows the layout of the start of an ADPCM audio frame.



Figure 5 - G.726 and IMA/DVI ADPCM Frame Metadata

The sequence number shall be incremented for every audio frame encoded, even if the frame was discarded (i.e. in noisy environments). The sequence number shall wrap back to 0×00 after $0 \times FF$.

January 20, 2020 Page 23 of 24

Both audio encoders shall use 4 bits per sample, and only one audio channel shall be sampled. The 4-bit samples are packed with the first sample in the four most significant bits and the second sample in the four least significant bits.

6.11.2.2 Audio Frame Buffering

Audio data is always buffered in 100 octet frames, and the device shall ensure that only complete frames are sent (as 5×20 octet GATT notifications). No partial frames shall ever be transferred. This ensures that the STB host can always determine where an audio frame starts and ends.

The RVS Server shall be able to buffer a minimum of 2 audio frames.

If frame buffers are exhausted then the complete frame shall be discarded, however the frame sequence number should still be incremented for the next encoded frame.

All buffered content shall be purged when either a disconnection event occurs, audio streaming is disabled via the Audio Control Characteristic or notifications are disabled for the characteristic.

January 20, 2020 Page 24 of 24