

vCon
Internet-Draft
Intended status: Standards Track
Expires: 29 March 2026

T. McCarthy-Howe
VCONIC
25 September 2025

vCon World Transcription Format Extension
draft-howe-vcon-wtf-extension-00

Abstract

This document defines the World Transcription Format (WTF) extension for Virtualized Conversations (vCon). The WTF extension provides a standardized method for representing speech-to-text transcription data from multiple providers within vCon containers. This extension enables consistent transcription storage, analysis, and interoperability across different transcription services while preserving provider-specific features through extensible fields.

The WTF extension is designed as a Compatible Extension that introduces transcription attachments without altering existing vCon semantics, ensuring backward compatibility with existing vCon implementations.

About This Document

This note is to be removed before publishing as an RFC.

The latest revision of this draft can be found at <https://vcon-dev.github.io/draft-howe-vcon-wtf-extension/draft-howe-vcon-wtf-extension-00.html>. Status information for this document may be found at <https://datatracker.ietf.org/doc/draft-howe-vcon-wtf-extension/>.

Discussion of this document takes place on the vCon Working Group mailing list (<mailto:vcon@ietf.org>), which is archived at <https://mailarchive.ietf.org/arch/browse/vcon/>. Subscribe at <https://www.ietf.org/mailman/listinfo/vcon/>.

Source for this draft and an issue tracker can be found at <https://github.com/vcon-dev/draft-howe-vcon-wtf-extension>.

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 <https://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 29 March 2026.

Copyright Notice

Copyright (c) 2025 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 (<https://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 Revised BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Revised BSD License.

Table of Contents

1. Requirements Language	3
2. Introduction	3
3. Conventions and Definitions	4
3.1. Core Terms	4
4. World Transcription Format Overview	5
4.1. Design Principles	5
4.2. Core Structure	5
5. vCon WTF Extension Definition	6
5.1. Extension Classification	6
5.2. Extension Registration	6
5.3. Extension Usage	6
6. WTF Attachment Structure	7
6.1. Attachment Container	7
6.2. WTF Body Structure	8
6.2.1. Required Fields	8
6.2.2. Optional Fields	9
7. Provider Integration Guidelines	11
7.1. Supported Providers	11
7.2. Conversion Requirements	11
7.3. Provider-Specific Mappings	11

7.3.1. Whisper Integration	11
7.3.2. Deepgram Integration	12
8. Quality and Confidence Metrics	12
8.1. Confidence Score Normalization	12
8.2. Quality Metrics	13
9. Security Considerations	13
9.1. Data Privacy	13
9.2. Provider-Specific Security	13
9.3. Integrity Protection	14
9.4. Temporal Validation	14
10. IANA Considerations	14
10.1. vCon Extensions Names Registry	14
10.2. WTF Attachment Type Values Registry	14
10.3. WTF Provider Registry	15
11. Examples	15
11.1. Basic Two-Party Call Transcription	15
11.2. Multi-Provider Transcription Comparison	18
12. References	20
12.1. Normative References	20
12.2. Informative References	20
Acknowledgements	21
Trademark Notice	21
Author's Address	21

1. Requirements Language

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.

2. Introduction

Virtualized Conversations (vCon) [I-D.draft-ietf-vcon-core-00] provide a standardized container format for conversation data, enabling interoperability across different communication platforms and modalities. As speech-to-text technology becomes increasingly important for conversation analysis, compliance, and accessibility, there is a growing need for standardized transcription representation within vCon containers.

Current transcription services each use proprietary formats, making it difficult to:

- * Switch between transcription providers without data conversion
- * Perform comparative analysis across different providers

- * Maintain consistent data processing pipelines
- * Preserve transcription metadata during system migrations
- * Ensure long-term accessibility and archival compliance

This document defines the World Transcription Format (WTF) extension for vCon, which addresses these challenges by providing:

- * A unified schema for transcription data from any speech-to-text provider
- * Hierarchical organization from words to segments to complete transcripts
- * Extensible fields for provider-specific features
- * Built-in quality metrics and confidence scoring
- * Support for real-time and batch transcription workflows
- * Consistent export capabilities to standard subtitle and caption formats

The WTF extension enables organizations to standardize their transcription workflows while maintaining the flexibility to use multiple providers and preserve provider-specific enhancements.

3. Conventions and Definitions

3.1. Core Terms

***World Transcription Format (WTF)*:** A standardized JSON schema for representing speech-to-text transcription data from any provider in a consistent, interoperable format.

***Transcription Provider*:** A service or system that converts audio or video content to text, such as Whisper(TM), Deepgram(TM), AssemblyAI(TM), Google Cloud Speech-to-Text(TM), Amazon Transcribe(TM), or Azure Speech Services(TM).

***WTF Attachment*:** A vCon attachment with type "wtf_transcription" that contains structured transcription information in World Transcription Format.

***Segment*:** A logical chunk of transcribed content, typically representing sentence or phrase boundaries with associated timing information.

***Speaker Diarization*:** The process of partitioning an audio stream into homogeneous segments according to the speaker identity, enabling "who spoke when" analysis.

***Compatible Extension*:** A vCon extension that introduces additional data without altering the meaning or structure of existing elements, as defined in [I-D.draft-ietf-vcon-core-00].

4. World Transcription Format Overview

The World Transcription Format uses a hierarchical JSON structure designed to capture transcription data at multiple levels of granularity while maintaining consistency across different providers.

4.1. Design Principles

***Completeness*:** WTF captures all essential transcription data including text, timing, confidence scores, and speaker information.

***Consistency*:** Provides uniform structure regardless of the underlying transcription provider.

***Extensibility*:** Supports provider-specific features through structured extension fields.

***Validation*:** Includes built-in data integrity checking and quality metrics.

***Hierarchical Organization*:** Natural progression from words to segments to complete transcripts.

4.2. Core Structure

WTF uses a hierarchical JSON structure with three required sections and multiple optional enrichment layers:

```
{
  "transcript": { /* Required: Core transcript information */ },
  "segments": [ /* Required: Time-aligned text segments */ ],
  "metadata": { /* Required: Processing metadata */ },
  "words": [ /* Optional: Word-level details */ ],
  "speakers": { /* Optional: Speaker diarization */ },
  "alternatives": [ /* Optional: Alternative transcriptions */ ],
  "enrichments": { /* Optional: Analysis features */ },
  "extensions": { /* Optional: Provider-specific data */ },
  "quality": { /* Optional: Quality metrics */ },
  "streaming": { /* Optional: Streaming information */ }
}
```

5. vCon WTF Extension Definition

5.1. Extension Classification

The WTF extension is a **Compatible Extension** as defined in Section 2.5 of [I-D.draft-ietf-vcon-core-00]. This extension:

- * Introduces transcription data without altering existing vCon semantics
- * Can be safely ignored by implementations that don't support transcription processing
- * Does not require listing in the `must_support` parameter
- * Maintains backward compatibility with existing vCon implementations

5.2. Extension Registration

This document defines the `"wtf_transcription"` extension token for registration in the vCon Extensions Names Registry:

- * **Extension Name**: `wtf_transcription`
- * **Extension Description**: World Transcription Format for standardized speech-to-text representation with multi-provider support
- * **Change Controller**: IESG
- * **Specification Document**: This document

5.3. Extension Usage

vCon instances that include WTF transcription attachments SHOULD include `"wtf_transcription"` in the extensions array:

```
{
  "uuid": "01234567-89ab-cdef-0123-456789abcdef",
  "extensions": ["wtf_transcription"],
  "created_at": "2025-01-02T12:00:00Z",
  "parties": [...],
  "dialog": [...],
  "attachments": [
    {
      "type": "wtf_transcription",
      "start": "2025-01-02T12:15:30Z",
      "party": 0,
      "dialog": 0,
      "encoding": "json",
      "body": {
        // WTF transcription data structure defined below
      }
    }
  ]
}
```

6. WTF Attachment Structure

6.1. Attachment Container

Transcription information **MUST** be included as vCon attachments using the standard attachment object structure defined in Section 4.4 of [I-D.draft-ietf-vcon-core-00].

The WTF transcription attachment **MUST** include:

- * ***type***: **MUST** be set to "wtf_transcription"
- * ***encoding***: **MUST** be set to "json" for structured transcription data
- * ***body***: **MUST** contain the WTF transcription data structure as defined below

The WTF transcription attachment **SHOULD** include:

- * ***start***: ISO 8601 timestamp when transcription was created
- * ***party***: Index of the party in the vCon parties array (for single-speaker transcription)
- * ***dialog***: Index of the associated dialog in the vCon dialog array

6.2. WTF Body Structure

The body field of the WTF transcription attachment MUST contain a JSON object with the following structure:

6.2.1. Required Fields

6.2.1.1. Transcript Object

```
"transcript": {  
  "text": "string",           // Complete transcription text  
  "language": "string",       // BCP-47 language code (e.g., "en-US")  
  "duration": "number",       // Total audio duration in seconds  
  "confidence": "number"      // Overall confidence score \[0.0-1.0\  
}
```

The transcript object provides the high-level summary of the entire transcription:

- * ***text***: The complete, concatenated transcription text
- * ***language***: MUST use BCP-47 format [BCP47] (examples: "en-US", "es-MX", "fr-CA")
- * ***duration***: Floating-point seconds, MUST be ≥ 0
- * ***confidence***: Normalized to [0, 1] range regardless of provider scale

6.2.1.2. Segments Array

```
"segments": [  
  {  
    "id": "integer",           // Sequential segment identifier  
    "start": "number",         // Start time in seconds  
    "end": "number",           // End time in seconds  
    "text": "string",          // Segment text content  
    "confidence": "number",     // Segment-level confidence \[0.0-1.0\  
    "speaker": "integer|string", // Optional: Speaker identifier  
    "words": ["integer"]        // Optional: Array of word indices  
  }  
]
```

Segments represent logical chunks of transcribed content, typically sentence or phrase boundaries:

- * ***id***: MUST be unique within the document, typically sequential

- * **start*/*end**: Floating-point seconds, where end > start
- * **text**: SHOULD be trimmed of leading/trailing whitespace
- * **speaker**: Can be integer (0, 1, 2) or string ("Speaker A")
- * **words**: References indices in the words array

6.2.1.3. Metadata Object

```
"metadata": {  
  "created_at": "string",           // ISO 8601 timestamp  
  "processed_at": "string",         // ISO 8601 timestamp  
  "provider": "string",             // Provider name (lowercase)  
  "model": "string",               // Model/version identifier  
  "processing_time": "number",      // Optional: Processing duration in seconds  
  "audio": {  
    "duration": "number",           // Source audio duration in seconds  
    "sample_rate": "integer",       // Optional: Sample rate in Hz  
    "channels": "integer",          // Optional: Number of channels  
    "format": "string",             // Optional: Audio format  
    "bitrate": "integer"           // Optional: Bitrate in kbps  
  },  
  "options": "object"              // Provider-specific options used  
}
```

The metadata object captures processing and source information:

- * **created_at*/*processed_at**: MUST use ISO 8601 format
- * **provider**: Lowercase identifier for supported providers
- * **model**: Provider's model identifier (e.g., "whisper-large-v3", "nova-2")
- * **options**: Preserves provider-specific configuration

6.2.2. Optional Fields

6.2.2.1. Words Array

```
"words": [  
  {  
    "id": "integer",           // Sequential word identifier  
    "start": "number",        // Word start time in seconds  
    "end": "number",          // Word end time in seconds  
    "text": "string",          // Word text  
    "confidence": "number",     // Word-level confidence \[0.0-1.0\  
    "speaker": "integer|string", // Optional: Speaker identifier  
    "is_punctuation": "boolean" // Optional: Punctuation marker  
  }  
]
```

6.2.2.2. Speakers Object

```
"speakers": {  
  "speaker_id": {  
    "id": "integer|string", // Speaker identifier  
    "label": "string",      // Human-readable speaker name  
    "segments": ["integer"], // Array of segment IDs for this speaker  
    "total_time": "number",  // Total speaking time in seconds  
    "confidence": "number"   // Diarization confidence \[0.0-1.0\  
  }  
}
```

6.2.2.3. Quality Object

```
"quality": {  
  "audio_quality": "string", // high, medium, low  
  "background_noise": "number", // Noise level \[0.0-1.0\  
  "multiple_speakers": "boolean",  
  "overlapping_speech": "boolean",  
  "silence_ratio": "number", // Percentage of silence  
  "average_confidence": "number",  
  "low_confidence_words": "integer",  
  "processing_warnings": ["string"]  
}
```

6.2.2.4. Extensions Object

```
"extensions": {  
  "provider_name": {  
    // Provider-specific fields preserved during conversion  
  }  
}
```

7. Provider Integration Guidelines

7.1. Supported Providers

The WTF extension supports integration with major transcription providers:

- * **Whisper(TM)**: OpenAI's open-source speech recognition system
- * **Deepgram(TM)**: Real-time speech-to-text API
- * **AssemblyAI(TM)**: AI-powered transcription and audio intelligence
- * **Google Cloud Speech-to-Text(TM)**: Google's speech recognition service
- * **Amazon Transcribe(TM)**: AWS speech-to-text service
- * **Azure Speech Services(TM)**: Microsoft's speech recognition platform
- * **Rev.ai(TM)**: Automated and human transcription services
- * **Speechmatics(TM)**: Real-time and batch speech recognition
- * **Wav2Vec2(TM)**: Facebook's self-supervised speech recognition model
- * **Parakeet(TM)**: NVIDIA's speech recognition toolkit

7.2. Conversion Requirements

When converting from provider-specific formats to WTF:

1. **Normalize confidence scores** to [0.0, 1.0] range
2. **Convert timestamps** to floating-point seconds
3. **Standardize language codes** to BCP-47 format
4. **Preserve provider-specific features** in extensions field
5. **Validate output** against WTF schema requirements

7.3. Provider-Specific Mappings

7.3.1. Whisper Integration

```
{
  "extensions": {
    "whisper": {
      "tokens": ["array of token IDs"],
      "temperature": "number",
      "compression_ratio": "number",
      "avg_logprob": "number",
      "no_speech_prob": "number"
    }
  }
}
```

7.3.2. Deepgram Integration

```
{
  "extensions": {
    "deepgram": {
      "utterances": ["array of utterance objects"],
      "paragraphs": ["array of paragraph objects"],
      "search_terms": ["array of detected search terms"]
    }
  }
}
```

8. Quality and Confidence Metrics

8.1. Confidence Score Normalization

All confidence scores MUST be normalized to the [0.0, 1.0] range:

- * *1.0*: Highest confidence (perfect recognition)
- * *0.9-1.0*: High confidence
- * *0.7-0.9*: Medium confidence
- * *0.5-0.7*: Low confidence
- * *0.0-0.5*: Very low confidence

Provider-specific scales are converted during import:

- * Percentage (0-100) -> divide by 100
- * Log probability -> exponential transformation
- * Custom scales -> linear normalization

8.2. Quality Metrics

The quality object provides assessment metrics:

- * `*audio_quality*`: Categorical assessment (high/medium/low)
- * `*background_noise*`: Noise level [0.0-1.0]
- * `*multiple_speakers*`: Boolean indicator of multi-speaker content
- * `*overlapping_speech*`: Boolean indicator of speaker overlap
- * `*silence_ratio*`: Percentage of audio that is silence
- * `*average_confidence*`: Mean confidence across all words/segments
- * `*low_confidence_words*`: Count of words below 0.5 confidence
- * `*processing_warnings*`: Array of processing issues or notices

9. Security Considerations

9.1. Data Privacy

Transcription data often contains sensitive personal information. Implementations SHOULD:

- * Apply appropriate access controls to WTF transcription attachments
- * Consider encryption requirements for transcription data at rest and in transit
- * Implement data retention policies consistent with privacy regulations
- * Provide mechanisms for transcription data redaction or anonymization

9.2. Provider-Specific Security

When integrating with external transcription providers:

- * Validate provider credentials and API security
- * Implement secure communication channels (TLS 1.2 or higher)
- * Consider data residency requirements for audio processing

- * Audit provider data handling practices and compliance certifications

9.3. Integrity Protection

WTF transcription attachments SHOULD be integrity protected using vCon signing mechanisms as defined in [I-D.draft-ietf-vcon-core-00] to prevent unauthorized modification of transcription data.

9.4. Temporal Validation

Implementations SHOULD validate that transcription timestamps are consistent with the associated dialog timing information to detect potential tampering or synchronization issues.

10. IANA Considerations

10.1. vCon Extensions Names Registry

This document requests IANA to register the following extension in the vCon Extensions Names Registry established by [I-D.draft-ietf-vcon-core-00]:

- * *Extension Name*: wtf_transcription
- * *Extension Description*: World Transcription Format for standardized speech-to-text representation with multi-provider support
- * *Change Controller*: IESG
- * *Specification Document*: This document

10.2. WTF Attachment Type Values Registry

This document requests IANA to establish a new registry for WTF attachment type values with the following initial registration:

- * *Type Value*: wtf_transcription
- * *Description*: Structured transcription records using World Transcription Format
- * *Change Controller*: IESG
- * *Specification Document*: This document Registration Template:
 - *Type Value*: The string value used as the attachment type identifier
 - *Description*: Brief description of the attachment type

and its purpose **Change Controller**: For Standards Track RFCs, list "IESG". For others, give the name of the responsible party. **Specification Document(s)**: Reference to defining documents with URIs where available

10.3. WTF Provider Registry

This document requests IANA to establish a new registry for WTF transcription providers with initial registrations for supported providers:

- * **Provider Name**: whisper
- * **Description**: OpenAI Whisper(TM) speech recognition system
- * **Change Controller**: IESG
- * **Specification Document**: This document (Additional provider registrations would be added for each supported provider)

11. Examples

11.1. Basic Two-Party Call Transcription

```
{
  "uuid": "01928e10-193e-8231-b9a2-279e0d16bc46",
  "vcon": "0.0.2",
  "extensions": ["wtf_transcription"],
  "created_at": "2025-01-02T12:00:00Z",
  "parties": [
    {
      "tel": "+1-555-123-4567",
      "name": "Alice"
    },
    {
      "tel": "+1-555-987-6543",
      "name": "Bob"
    }
  ],
  "dialog": [
    {
      "type": "recording",
      "start": "2025-01-02T12:15:30Z",
      "duration": 65.2,
      "parties": [0, 1],
      "mediatype": "audio/wav",
      "filename": "call-recording.wav"
    }
  ]
}
```

```
],
"attachments": [
  {
    "type": "wtf_transcription",
    "start": "2025-01-02T12:16:35Z",
    "dialog": 0,
    "encoding": "json",
    "body": {
      "transcript": {
        "text": "Hello, this is Alice from customer service. " +
              "How can I help you today? Hi Alice, I'm having " +
              "trouble with my account. Can you help me reset " +
              "my password?",
        "language": "en-US",
        "duration": 65.2,
        "confidence": 0.92
      },
      "segments": [
        {
          "id": 0,
          "start": 0.5,
          "end": 4.8,
          "text": "Hello, this is Alice from customer service. " +
                "How can I help you today?",
          "confidence": 0.95,
          "speaker": 0,
          "words": [0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14]
        },
        {
          "id": 1,
          "start": 5.2,
          "end": 9.1,
          "text": "Hi Alice, I'm having trouble with my account. " +
                "Can you help me reset my password?",
          "confidence": 0.88,
          "speaker": 1,
          "words": [16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29,
                  30, 31]
        }
      ]
    },
    "words": [
      {
        "id": 0,
        "start": 0.5,
        "end": 0.8,
        "text": "Hello",
        "confidence": 0.98,
        "speaker": 0,
```

```
    "is_punctuation": false
  },
  {
    "id": 1,
    "start": 0.9,
    "end": 1.1,
    "text": ",",
    "confidence": 0.95,
    "speaker": 0,
    "is_punctuation": true
  }
  // Additional words...
],
"speakers": {
  "0": {
    "id": 0,
    "label": "Alice (Customer Service)",
    "segments": [0],
    "total_time": 4.3,
    "confidence": 0.95
  },
  "1": {
    "id": 1,
    "label": "Bob (Customer)",
    "segments": [1],
    "total_time": 3.9,
    "confidence": 0.88
  }
},
"metadata": {
  "created_at": "2025-01-02T12:15:30Z",
  "processed_at": "2025-01-02T12:16:35Z",
  "provider": "deepgram",
  "model": "nova-2",
  "processing_time": 12.5,
  "audio": {
    "duration": 65.2,
    "sample_rate": 8000,
    "channels": 1,
    "format": "wav",
    "bitrate": 128
  },
  "options": {
    "punctuate": true,
    "diarize": true,
    "language": "en"
  }
},
},
```

```

    "quality": {
      "audio_quality": "high",
      "background_noise": 0.1,
      "multiple_speakers": true,
      "overlapping_speech": false,
      "silence_ratio": 0.15,
      "average_confidence": 0.92,
      "low_confidence_words": 0,
      "processing_warnings": []
    },
    "extensions": {
      "deepgram": {
        "utterances": [
          {
            "start": 0.5,
            "end": 4.8,
            "confidence": 0.95,
            "channel": 0,
            "transcript": "Hello, this is Alice from customer service. " +
                          "How can I help you today?"
          }
        ]
      }
    }
  }
}

```

11.2. Multi-Provider Transcription Comparison

```

{
  "attachments": [
    {
      "type": "wtf_transcription",
      "dialog": 0,
      "encoding": "json",
      "body": {
        "transcript": {
          "text": "The quick brown fox jumps over the lazy dog.",
          "language": "en-US",
          "duration": 3.5,
          "confidence": 0.96
        },
        "segments": [
          {
            "id": 0,
            "start": 0.0,

```

```
        "end": 3.5,
        "text": "The quick brown fox jumps over the lazy dog.",
        "confidence": 0.96,
        "speaker": 0
    }
],
"metadata": {
    "created_at": "2025-01-02T12:00:00Z",
    "processed_at": "2025-01-02T12:00:15Z",
    "provider": "whisper",
    "model": "whisper-large-v3",
    "processing_time": 15.2
},
"extensions": {
    "whisper": {
        "temperature": 0.0,
        "compression_ratio": 2.1,
        "avg_logprob": -0.25,
        "no_speech_prob": 0.01
    }
}
},
{
    "type": "wtf_transcription",
    "dialog": 0,
    "encoding": "json",
    "body": {
        "transcript": {
            "text": "The quick brown fox jumps over the lazy dog.",
            "language": "en-US",
            "duration": 3.5,
            "confidence": 0.94
        },
        "segments": [
            {
                "id": 0,
                "start": 0.0,
                "end": 3.5,
                "text": "The quick brown fox jumps over the lazy dog.",
                "confidence": 0.94,
                "speaker": 0
            }
        ],
        "metadata": {
            "created_at": "2025-01-02T12:00:00Z",
            "processed_at": "2025-01-02T12:00:08Z",
            "provider": "deepgram",
```

```
    "model": "nova-2",
    "processing_time": 8.1
  },
  "extensions": {
    "deepgram": {
      "model_info": {
        "name": "nova-2",
        "version": "2024-01-09",
        "uuid": "4d892fb6-7cc1-4e7a-a1b3-1c2e3f4a5b6c"
      }
    }
  }
}
```

12. References

12.1. Normative References

- [I-D.draft-ietf-vcon-core-00]
Petrie, D., "Virtualized Conversation (vCon) Container",
Work in Progress, Internet-Draft, draft-ietf-vcon-core-00,
March 2025, <I-D.draft-ietf-vcon-core-00>.
- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate
Requirement Levels", BCP 14, RFC 2119,
DOI 10.17487/RFC2119, March 1997,
<<https://www.rfc-editor.org/rfc/rfc2119>>.
- [RFC3339] Klyne, G., "Date and Time on the Internet: Timestamps",
July 2002, <<https://www.rfc-editor.org/rfc/rfc3339.html>>.
- [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/rfc/rfc8174>>.
- [RFC8949] Bormann, C., "Concise Binary Object Representation
(CBOR)", December 2020,
<<https://www.rfc-editor.org/rfc/rfc8949.html>>.

12.2. Informative References

- [BCP47] Phillips, A., "Tags for Identifying Languages", September
2009, <<https://www.rfc-editor.org/rfc/rfc5646.html>>.

[I-D.draft-ietf-vcon-overview]

McCarthy-Howe, T., "The vCon - Conversation Data Container - Overview", Work in Progress, Internet-Draft, draft-ietf-vcon-overview-00, July 2025, <I-D.draft-ietf-vcon-overview-00>.

[WHISPER] Radford, A., Kim, J. W., Xu, T., Brockman, G., McLeavey, C., and I. Sutskever, "Whisper: Robust Speech Recognition via Large-Scale Weak Supervision", September 2022, <<https://openai.com/research/whisper>>.

Acknowledgements

- * Appreciation to the transcription provider community for their input on standardization requirements.
- * Thanks to the vCon working group for their feedback and guidance on extension design patterns.

Trademark Notice

All trademarks mentioned in this document are the property of their respective owners. The use of these trademarks does not imply endorsement by the IETF or the authors of this document. The following trademarks are referenced:

- * Whisper(TM) is a trademark of OpenAI, Inc.
- * Deepgram(TM) is a trademark of Deepgram, Inc.
- * AssemblyAI(TM) is a trademark of AssemblyAI, Inc.
- * Google Cloud Speech-to-Text(TM) is a trademark of Google LLC.
- * Amazon Transcribe(TM) is a trademark of Amazon.com, Inc.
- * Azure Speech Services(TM) is a trademark of Microsoft Corporation.
- * Rev.ai(TM) is a trademark of Rev.com, Inc.
- * Speechmatics(TM) is a trademark of Speechmatics Limited.
- * Wav2Vec2(TM) is a trademark of Meta Platforms, Inc.
- * Parakeet(TM) is a trademark of NVIDIA Corporation.

Author's Address

Thomas McCarthy-Howe
VCONIC
Email: ghostofbasho@gmail.com