AVX

Acoustic Signal Vocabulary for Timed Text

Version 1.0 Specification

Published by:	Cognispace LLC
Date:	May 2026
Status:	Version 1.0 — Initial Publication
Namespace URI:	`https://cognispacehq.com/auravoxa/caption/1.0`

This document is the normative specification for the AVX 1.0 vocabulary. It is published at the canonical namespace URI above.

Table of Contents

1. Abstract
2. Status of This Document
3. Namespace Declaration
4. Conformance
5. Governance Envelope
6. Signal Vocabulary
  6.1 Utterance-Level Attributes
  6.2 Span-Level Attributes
7. Complete Example
8. Rendering Guidance (Non-Normative)
9. Out of Scope
10. AVX Specification License
11. Version History

1. Abstract

AVX is a metadata vocabulary for annotating timed text with acoustic speech signal observations. It is designed as a foreign namespace extension for TTML2 (W3C Recommendation, 2018) and as a JSON sidecar format for WebVTT. It is backward-compatible with all conformant TTML2 parsers. It is versioned. It is published and maintained by Cognispace LLC. This document specifies version 1.0 of the AVX vocabulary.

2. Status of This Document

This document specifies AVX version 1.0, published May 2026. Cognispace LLC is the maintaining authority for this specification. This version supersedes no prior version. Future versions will be published at incrementing version paths under the same base URI.

This specification is published under the AVX Specification License (see Section 10). Implementation is permitted under that license. The AVX-Compliant certification mark is a separate trademark of Cognispace LLC and is not granted by this license.

3. Namespace Declaration

The canonical namespace URI:

```
xmlns:avx="https://cognispacehq.com/auravoxa/caption/1.0"
```

> *This URI is the normative identifier for the AVX 1.0 namespace. It is not required to resolve to a web page for the namespace to be valid in XML processing — however, this specification is published at this URI as a convenience to implementers.*

> *The avx: prefix is the recommended prefix for this namespace. Implementations may use any prefix; the namespace URI is the authoritative identifier.*

4. Conformance

AVX Document

> *A conformant AVX Document is a TTML2 document that uses the AVX namespace and whose avx: attributes conform to the vocabulary defined in this specification. A conformant AVX Document must include the governance envelope attributes defined in Section 5.*

AVX Renderer

> *A conformant AVX Renderer is software that reads AVX attributes and produces enriched output according to those attributes. A renderer that does not recognize avx: namespace attributes must ignore them silently*

and render the baseline tts: content normally. This requirement is not advisory — silent graceful degradation is a conformance requirement of this specification.

5. Governance Envelope

These attributes appear on the root <tt> element. They are required in all conformant AVX Documents.

Attribute	Type	Required	Description
avx:version	string	Required	AVX vocabulary version. Current value: "1.0"
avx:governance	string	Required	Governing specification. Value: "AVX-1.0" — refers to this document
avx:claim_boundary	string	Required	Observation boundary declaration. The only permitted value in version 1.0 is "observational"
avx:context_type	enum	Required	Content context. Values: live_unscripted, scripted_fiction, professional_review, educational

On avx:claim_boundary

The value "observational" is the only permitted value in this version of the specification. It declares that all signal values in the document are observations of acoustic properties of speech — measurements of how words were spoken — and are not claims about mental state, emotional state, psychological condition, deception, intent, or any other construct that cannot be directly measured from acoustic signal data alone.

6. Signal Vocabulary

6.1 Utterance-Level Attributes

These attributes appear on TTML <p> elements. They describe properties of the utterance as a whole.

Attribute	Type	Values	Acoustic Basis	Description
avx:arc	enum	ASCENDING, SUSTAINED, RESOLVED, TAPERING, CARRIED, DROPPED	Trajectory of acoustic energy and pitch movement across the utterance	Direction of acoustic energy across the utterance. ASCENDING: energy builds toward the end. SUSTAINED: flat throughout. RESOLVED: rises then recovers. TAPERING: controlled descent. CARRIED: energy continues into next utterance. DROPPED: collapses abruptly.

avx:coherence	float	0.0–1.0	Estimated alignment between acoustic delivery and lexical content	Estimated alignment between how the utterance is spoken and what is said. Higher values indicate higher alignment. This is an observational estimate — not a claim about intent or truthfulness.
avx:effort	enum	low, medium, high	Inferred acoustic effort derived from prosodic emphasis, voicing continuity, energy variance, and trajectory	Inferred pattern of acoustic effort in this utterance. This is an acoustic inference — not a claim about the speaker's mental or emotional state.
avx:emphasis	enum	suppressed, distributed, concentrated, sparse	Distribution pattern of prosodic emphasis across the words in the utterance	suppressed: emphasis absent despite high voicing. distributed: spread across multiple words. concentrated: focused on one or two words. sparse: present but minimal.

6.2 Span-Level Attributes

These attributes appear on TTML elements. They describe properties of individual words or word groups.

Attribute	Type	Values	Acoustic Basis	Description
avx:eld	enum	low, medium, high, highest	Relative acoustic energy for this word compared to surrounding words in the utterance	Expressive energy for this word relative to other words in the utterance. Reflects acoustic prominence.
avx:arousal	float	0.0–1.0	Fundamental frequency and onset energy of this word	Acoustic activation level. Higher values reflect higher acoustic activation — more tonal and percussive energy.
avx:voiced	float	0.0–1.0	Proportion of this word's duration in which the speaker's vocal folds were fully vibrating	Proportion of fully voiced delivery. Higher values indicate fully voiced speech. Lower values indicate breathed, whispered, or suppressed delivery.
avx:hedge	boolean	true, false	Presence of hedge markers detected through lexical analysis of the utterance	Whether this word introduces epistemic uncertainty, tentativeness, or qualification into the utterance.

7. Complete Example

A full TTML2+AVX document showing all document-level, utterance-level, and span-level attributes in use.

```
<!-- tts: attributes = standard TTML styling layer (any player) -->
<!-- avx: attributes = acoustic signal layer (AVX-aware renderers) -->
<?xml version="1.0" encoding="UTF-8"?>
<tt xml:lang="en"
    xmlns="http://www.w3.org/ns/ttml"
    xmlns:tts="http://www.w3.org/ns/ttml#styling"
    xmlns:avx="https://cognispacehq.com/auravoxa/caption/1.0"
    avx:version="1.0"
    avx:governance="AVX-1.0"
    avx:claim_boundary="observational"
    avx:context_type="live_unscripted">
  <head>
    <styling>
      <style xml:id="s1" tts:fontFamily="Arial" tts:fontSize="100%"
        tts:color="white" tts:backgroundColor="rgba(0,0,0,0.82)"
        tts:padding="4px 10px"/>
    </styling>
  </head>
  <body><div>
    <!-- Utterance 1: tapering trajectory, high effort, suppressed emphasis -->
    <p xml:id="utt-01" begin="00:00:01.20" end="00:00:03.80" style="s1"
       avx:arc="TAPERING" avx:coherence="0.31"
       avx:effort="high" avx:emphasis="suppressed">
      <span tts:fontWeight="normal"
            avx:eld="low" avx:arousal="0.28" avx:voiced="0.91">I </span>
      <span tts:fontWeight="bold"
            avx:eld="high" avx:arousal="0.87" avx:voiced="0.22">know </span>
      <span tts:fontWeight="bold"
            avx:eld="highest" avx:arousal="0.92" avx:voiced="0.19">there.</span>
    </p>
    <!-- Utterance 2: sustained trajectory, hedge present -->
    <p xml:id="utt-02" begin="00:00:04.50" end="00:00:08.10" style="s1"
       avx:arc="SUSTAINED" avx:coherence="0.38"
       avx:effort="high" avx:emphasis="suppressed">
      <span tts:fontWeight="normal" tts:fontStyle="italic"
            avx:eld="medium" avx:arousal="0.47" avx:voiced="0.87"
            avx:hedge="true">possibly </span>
      <span tts:fontWeight="bold"
            avx:eld="highest" avx:arousal="0.93" avx:voiced="0.95"
            avx:hedge="false">significant </span>
    </p>
    <!-- Utterance 3: ascending trajectory, fully voiced -->
    <p xml:id="utt-03" begin="00:00:09.00" end="00:00:12.40" style="s1"
       avx:arc="ASCENDING" avx:coherence="0.78"
       avx:effort="low" avx:emphasis="concentrated">
      <span tts:fontWeight="bold"
            avx:eld="highest" avx:arousal="0.96" avx:voiced="0.98"
            avx:hedge="false">case </span>
      <span tts:fontWeight="bold"
            avx:eld="high" avx:arousal="0.88" avx:voiced="0.95"
            avx:hedge="false">dismissed.</span>
    </p>
  </div></body>
</tt>
```

8. Rendering Guidance (Non-Normative)

> *This section is non-normative. It describes how AVX attributes map to visual rendering decisions in reference implementations. Implementations are not required to follow these mappings exactly, but must conform to the governance constraints in Sections 4 and 9.*

Word weight mapping (avx:eld)

avx:eld	Recommended font-weight	Recommended font-size
low	400	95%
medium	500	100%
high	600	108%
highest	700	116%

Color temperature mapping (avx:voiced)

voiced range	Temperature	Example color
< 0.35	Cool	#B8D8DC (blue-teal)
0.35 – 0.70	Neutral	#E8DDD0 (warm white)
> 0.70	Warm	#F5C4A0 (peach)

Italic mapping (avx:hedge)

avx:hedge	Rendering
true	italic
false	roman

9. Out of Scope

- AVX does not define how signal values are computed. The acoustic measurements that produce AVX attribute values are not specified by this vocabulary.
- AVX does not define or constrain rendering behavior beyond the conformance requirement that unknown attributes be ignored silently.
- AVX does not detect, identify, or classify emotional states.
- AVX does not detect, identify, or assess mental states of any kind.
- AVX does not assess deception, dishonesty, evasion, or intent to mislead.
- AVX does not assess the truthfulness or credibility of speech.
- AVX does not assess the psychological condition of speakers.
- AVX does not provide clinical, therapeutic, diagnostic, or forensic outputs.

- AVX attribute values are observations of acoustic properties. Cause is not determinable from acoustic evidence alone.
- Renderers must not use AVX values to display indicators that imply judgment, scoring, ranking, or behavioral assessment of speakers.

10. AVX Specification License

Copyright 2026 Cognispace LLC. All rights reserved.

Permission is hereby granted, free of charge, to any person or organization to read, implement, and distribute implementations conforming to this specification, subject to the following conditions:

1. Any implementation claiming conformance to this specification must conform to all normative requirements in this document.
2. This license does not grant permission to publish modified versions of this specification that claim to be AVX version 1.0. Modified versions must be clearly identified as derivative works and must not use the name "AVX" without a distinguishing qualifier.
3. The AVX-Compliant certification mark is a trademark of Cognispace LLC. Use of the certification mark requires a separate trademark license. This specification license does not grant use of the certification mark.
4. This license does not grant any rights under any Cognispace LLC patents.

THE SPECIFICATION IS PROVIDED "AS IS" WITHOUT WARRANTY OF ANY KIND.

11. Version History

Version	Date	Summary
1.0	May 2026	Initial publication. Governance envelope, utterance-level attributes (avx:arc, avx:coherence, avx:effort, avx:emphasis), span-level attributes (avx:eld, avx:arousal, avx:voiced, avx:hedge), conformance classes, and AVX Specification License.