Transcription and Subtitle Processing Lambda – Functional guide

Hosting Infrastructure · https://www.mooplugins.com/docs/video-lesson-activity/hosting-infrastructure/transcription-and-subtitle-processing-lambda-functional-guide/

Purpose of this Lambda

The Transcription and Subtitle Processing Lambda is responsible for **audio transcription and subtitle generation** for processed media used by Video Lesson Activity.

It manages:

- Speech-to-text transcription
- WebVTT subtitle generation
- Optional subtitle translation into multiple languages
- On-demand subtitle requests
- Persistent subtitle discovery and state tracking

This Lambda operates **independently** from video transcoding and does not block video availability.

Role in the overall architecture

This Lambda is activated **after a processed media file becomes available**.

Its role is to:

- Convert spoken audio into structured transcript data
- Generate readable, synchronized subtitle files
- Fulfill subtitle requests regardless of timing
- Maintain consistent subtitle state for external systems such as Moodle

It ensures subtitles are generated **reliably**, even when requests occur before transcription has completed.

High-level workflow

At a functional level, the workflow follows this sequence:

1. Processed media output becomes available
2. Transcription is requested or triggered
3. Speech-to-text processing runs
4. Transcript data is stored
5. Subtitles are generated from transcript data
6. Optional translation is applied
7. WebVTT subtitle files are persisted
8. Available subtitles are exposed for discovery

Each step is event-driven and resilient to retries.

Event sources

The Lambda responds to three categories of events.

API-based requests

Used to:

- Request subtitle generation for a specific language
- Trigger transcription or subtitle creation on demand
- Act as an integration point for external systems such as LMS user interfaces

API requests allow subtitles to be requested **at any point**, even if transcription has not yet started.

Notification events

Used to:

- Start transcription for newly processed media
- Request additional subtitle languages

- Decouple subtitle processing from other workflows

Each notification includes:

- A clear action (transcribe or subtitle)
- Context required to locate media and store outputs

Transcription job lifecycle events

Triggered when the transcription service updates job status.

Used to:

- Detect transcription completion
- Retrieve transcript data
- Trigger subtitle generation and post-processing

Lifecycle handling is idempotent and tolerant of repeated events.

Transcription job creation

When transcription is requested:

- The Lambda verifies that the media file exists in storage
- A transcription job is created with:
  - Automatic language identification
  - Predictable output locations
- Transcript data is written to a known subtitles directory

Retries are applied to handle eventual consistency when media files are newly generated.

Transcript handling

After transcription completes:

- Transcript JSON is retrieved from storage
- The dominant language is detected
- Word-level timing information is extracted

This transcript becomes the **single source of truth** for all subtitle generation and translation.

Subtitle segmentation

Transcript data is converted into readable subtitle segments based on:

- Maximum word count per segment
- Maximum character length
- Natural punctuation boundaries

Each segment includes:

- Start timestamp
- End timestamp
- Text content

This segmentation ensures subtitles are:

- Easy to read
- Properly synchronized
- Suitable for long-form learning content

Language translation

For subtitle languages other than the original:

- Each subtitle segment is translated independently
- Translation runs concurrently for efficiency
- Original timing information is preserved

Language handling uses standardized language codes to ensure consistent naming and discovery.

WebVTT generation

Subtitles are generated in **WebVTT format**, which is widely supported by video players.

During generation:

- Timestamps are formatted to WebVTT specifications

- Each subtitle block includes translated or original text
- Files are stored using a consistent structure

Typical storage pattern:

```
/subtitles/{language}.vtt
```

These files are immediately usable for playback.

On-demand subtitle requests

Subtitles can be requested at any time.

When a subtitle request is received:

- If the subtitle already exists, the request completes immediately
- If a transcript exists, the subtitle is generated immediately
- If no transcript exists:
  - A pending language marker is created
  - Transcription is started automatically

This ensures subtitle requests are **never lost**, regardless of timing.

Pending subtitle handling

Pending subtitle requests are tracked using lightweight marker files.

Example structure:

```
/subtitles/pending/{language}.pending
```

When transcription completes:

- All pending languages are discovered
- Subtitles are generated for each language
- Pending markers are removed

This mechanism guarantees eventual subtitle fulfillment without manual retries.

Subtitle discovery

The Lambda supports subtitle discovery by:

- Scanning the subtitles directory
- Mapping language codes to human-readable names
- Returning structured metadata to external systems

This allows Moodle to:

- Display available subtitle options dynamically
- Avoid hard-coded language lists
- Reflect subtitle availability accurately

Error handling and retries

The Lambda includes safeguards for:

- Missing or delayed media files
- Temporary storage access failures
- Transcription service delays
- Duplicate or repeated events

Idempotent checks and controlled retries ensure stable behavior under load.

Security and access model

The Lambda operates with scoped permissions limited to:

- Reading and writing transcript and subtitle storage paths
- Creating and monitoring transcription jobs
- Language detection and translation services

All access is role-based and restricted to required resources.

Design characteristics

This Lambda is designed to be:

- Event-driven
- Decoupled from video transcoding
- Supportive of multiple subtitle languages
- Resilient to delayed dependencies
- Efficient through concurrent processing
- Easy to integrate with LMS platforms

Implementation summary

At a functional level, the Transcription and Subtitle Processing Lambda:

- Accepts transcription and subtitle requests via events or API
- Creates transcription jobs when media becomes available
- Generates readable, synchronized subtitles
- Translates subtitles when requested
- Tracks and fulfills pending subtitle requests
- Stores subtitles in a predictable, discoverable structure