

Detailed Use Case Description: ASR Training Within Customer Boundary

Use Case

Axon's Automatic Speech Recognition (ASR) technology converts body-worn camera and other evidence audio into written transcripts that officers review and, when needed, correct before submitting them as legal evidence. This use case describes a privacy-protected process that uses those officer corrections, together with short matching audio clips, to improve Axon's speech-to-text models entirely within the customer's secure environment.

Improving speech-to-text for real body-camera audio requires learning from accurate, real-world examples. Longer-term programs that carefully prepare and annotate recordings remain an important path for building large training datasets. This use case complements those programs by using corrections officers already make in the normal course of work — without asking customers for a separate annotation project, without Axon personnel reviewing transcripts or listening to audio, and without moving raw audio or raw transcript text outside the customer environment for this training process.

This process is eyes-off and ears-off. Eyes-off means no Axon individual views customer transcript text. Ears-off means no Axon individual listens to customer evidence audio. Only automated systems process that content inside the customer environment.

Customer Benefit

This use case helps Axon improve speech-to-text quality in the real conditions where transcription is hardest — for example noisy body-camera audio — by learning from corrections officers already make. Customers benefit from progress toward more accurate transcripts while remaining eyes-off and ears-off: raw evidence audio and full transcript content stay inside the customer environment for this process, are never reviewed or listened to by Axon personnel, and do not require a separate labeling project for each improvement cycle.

Data Accessed

For eligible use of this feature, Axon's automated systems may access the following data within the customer environment. No Axon individual reviews transcript text or listens to evidence audio as part of this use case:

Records that show when a user saved a modified transcript, including limited identifiers needed to locate the related evidence.

Transcript content for those evidences, accessed eyes-off by automation only, used solely to find word-level corrections and to select short portions of the transcript that contain those corrections. No individual reads this transcript content.

Evidence audio for evidences that still have eligible short portions after correction screening and personal-information filtering, accessed by automation only, used solely to cut short audio clips and measure basic audio quality (for example how noisy the clip is). No individual listens to this audio.

Separate evaluation materials that Axon places into the customer environment only to measure whether an improved model is more accurate. Those materials are not the customer's production evidence.

How Axon Uses the Data

Axon uses this information to run an improvement process entirely within the customer environment. At every stage below, customer transcript text and customer evidence audio are handled only by automation; Axon personnel do not view or listen to that content:

- **Finding eligible examples.** Starting from evidences where a transcript was saved after edits, automated systems look for word-level corrections, select short speaker turns that contain at least one corrected word and meet length limits, discard any portion where personal information is detected (or if that check cannot be completed), and calculate simple quality numbers such as how noisy the audio is and how confident the original transcription was. No individual inspects transcript text or audio during this step. The result is (1) a restricted internal record used later by automation for training and (2) a de-identified summary of numbers only — with no organization or evidence identifiers, no transcript text, and no audio — for Axon personnel to review.
- **Setting filters.** Axon personnel review only those de-identified numbers (for example audio noisiness, transcription confidence, clip length, word counts, and how many words were corrected) and choose simple limits such as how noisy a clip may be, how many corrections a clip must contain, allowed word-count ranges, and a maximum confidence level. Personnel do not see organization or evidence identifiers, do not see transcript text, and do not listen to audio.
- **Training and checking the model.** Automation applies the approved limits to select clips, briefly retrieves the matching audio inside the customer environment, stores those short clips only as long as needed with no human access, and uses each pair of short audio clip and corrected text to improve the speech-to-text model*. The improved model is checked against Axon's evaluation materials inside the customer environment. The updated model and accuracy results from the evaluation materials may be shared with Axon personnel for review; those results do not include customer transcript text or customer evidence audio.
- ***Addendum:** As of 9/28, those same selected clips are also used for a second training approach based on the original transcript text. Comparing that approach with training on the officer-reviewed transcript helps determine whether officer corrections improve model accuracy.

Privacy-Preserving Technique

- All processing for this use case runs within the customer's secure Axon environment under ACEIP on an eyes-off, ears-off basis. Raw customer transcripts and evidence audio are accessed only by automated systems during discovery and training. No Axon individual views raw transcript content or listens to audio at any.
- Before any short clip may be used for training, its text is automatically checked for personal information such as names, phone numbers, social security numbers, dates of birth, email addresses, physical addresses, badge numbers, case numbers, license plates, vehicle identification numbers, and other sensitive identifiers. If any personal information is detected the clip is discarded and is not used for training. Only clips that pass this check remain eligible.

- Training clips are further limited to short portions: no more than 40 seconds and no more than 50 words. These limits reduce how much content can enter training and further reduce the chance that sensitive content is used.
- Short audio and text clips used for training are kept only briefly (for example up to 24 hours*), with no human access, are used only by automated training, and are automatically deleted afterward whether training succeeds or not.
- ***Addendum:** As of 9/21, we are increasing the duration of retention to 72 hours.
- What may leave the customer environment for this use case is limited to (1) the de-identified numbers used to choose filter limits and (2) evaluation results on Axon's own evaluation materials (for example overall accuracy scores).

Data Schema

The following categories of data are relevant to this use case. Only the de-identified numbers and evaluation results described below may leave the customer environment for Axon personnel review.

Restricted internal record used by automation (inside the customer environment; not accessible to individuals):

- **Organization ID:** an identifier for the customer organization associated with the evidence.
- **Evidence ID:** an identifier for the evidence recording.
- **Average transcription confidence:** a number reflecting how confident the original transcription was.
- **Number of speakers:** how many speakers appear in the transcript.
- **Number of words edited:** how many words were corrected by a person.
- **Number of segments:** how many short eligible portions contain corrections.
- **Words per minute:** a speaking-rate number for the evidence.
- **Segment details:**
 - start and end times
 - duration
 - word count
 - number of corrected words
 - the numbered index of the corrected word in this segment
 - a speaker label such as Speaker 0 or Speaker 1, not the actual speaker identity
 - an audio-quality number (how noisy the clip is)
 - average confidence score from the current production ASR model
 - the words in the clip with confidence score, correction flags and timing.

De-identified summary for Axon personnel (leave the customer environment; no identifiers, no transcript text, no audio):

- **Average transcription confidence:** a confidence number from the current production ASR model
- **Number of speakers:** a speaker-count number.
- **Number of words edited:** a correction-count number.
- **Number of segments:** a segment-count number.

- **Words per minute:** a speaking-rate number.
- **Segment statistics:** length, word count, number of corrected words, edit type, audio-quality number, and confidence scores of the corrected words

Temporary training clips (inside the customer environment; short retention; not accessible to individuals):

- **Audio clip:** a short audio portion cut to the approved start and end times.
- **Corrected transcript text:** the matching short text used to teach the model after personal-information filtering.
- **Addendum:** As of 9/28, the **original transcript text** for the same short clip is also kept for the second training approach, under the same personal-information filtering and the same short retention.

Training and evaluation outputs:

- **Updated model:** the improved speech-to-text model weights, these are pure numbers with no-identifiable content.
- **Evaluation results:** accuracy and related scores computed on Axon's evaluation materials; these may be shared with Axon personnel and do not include customer transcript text or customer evidence audio.

Examples of What May or May Not Be Extracted

See example below:

Original Customer Data	De-identified Numbers for Human Review
<pre>{ "organization_id": "550e8400-e29b-41d4-a716-446655440000", "evidence_id": "123e4567-e89b-12d3-a456-426614174000", "average_confidence": 0.82, "number_of_speakers": 2, "number_of_words_edited": 7, "number_of_segments": 3, "words_per_minute": 142, "segments": [{ "segment_id": "0", "start_time_seconds": 1.51, "end_time_seconds": 14.8, "duration_seconds": 13.29, "word_count": 18, "edited_word_index": [2, 3], "speaker_label": "0", "audio_quality": 15.3, "average_confidence": 0.71,</pre>	<pre>{ "average_confidence": 0.82, "number_of_speakers": 2, "number_of_words_edited": 7, "number_of_segments": 3, "words_per_minute": 142, "segments": [{ "segment_id": "0", "start_time_seconds": 1.51, "end_time_seconds": 14.8, "duration_seconds": 13.29, "word_count": 18, "edited_word_index": [2, 3], "audio_quality": 15.3, "average_confidence": 0.71, "modified_words_confidence": [0.82, 0.61], }] }</pre>

<pre> "modified_words_confidence":[0.82, 0.61], "words": [{ "text": "while", "corrected": true, "start_time_seconds": 5.5, "end_time_seconds": 5.6 }, { "text": "the", "corrected": false, "start_time_seconds": 5.7, "end_time_seconds": 5.8 }] }] } </pre>	
---	--

Segments where Person Identifiable Information presents will not be exported or used in training.

Segments Will Be Used for Training	Segment Will Not Be Used for Training
Alright, we're doing quick breath checks tonight.	name is Emily Carter.
I haven't had anything to drink or taken any medication, if you need to know.	For example, you can reach the shelter at 555-402-7891 or the outreach team at (555) 336-1402.
Where were you coming from before you went into the store	If you keep refusing, I'll have to call for backup and escalate this, Officer Bennett will be here shortly.

Retention Period

Restricted internal records used by automation are kept only as long as needed to select examples and run training (up to 30 days so enough eligible examples can accumulate and filter limits can be set before training runs), then automatically deleted. Those records remain eyes-off and ears-off for their entire lifetime — no individual access. Temporary audio and text clips used for training are kept only for the short period needed to finish a training run (for example up to 24 hours*), likewise with no human access, and are automatically deleted afterward whether training succeeds or not. ***Addendum:** As of 9/21, we are increasing the duration of retention to 72 hours. De-identified numbers and evaluation results from Axon's evaluation materials may be kept for a limited period no longer than necessary to support model improvement for this use case. Customer evidence audio and customer transcript text are not kept outside the customer environment for this process and are never kept in a form available for human viewing or listening.

Preservation of Original Content

This use case does not modify the customer's original evidence, source transcripts, or live product behavior. Original evidence and transcripts in customer systems remain unchanged. Automated systems read transcript and audio content temporarily within the customer environment for finding examples and training; this use case does not change how transcription works for officers in production. Turning on an improved model for end users is outside the scope of this description.

Withdrawal

Customers participating in this use case may request withdrawal. After withdrawal, Axon would stop collecting new data for this use case from that customer and would stop including that customer's evidences in this improvement process. Previously generated de-identified numbers, evaluation results, and model improvements that no longer identify the customer, user, or original evidence may continue to be retained where permitted.

Summary

This use case helps Axon improve ASR models by training on short clips that contain officer word-level transcript corrections and that have passed an automated personal-information check, entirely within the customer environment under an eyes-off, ears-off model. Automated systems access transcript and audio content only as needed to find corrections, select short clips, discard clips with personal information, calculate quality numbers, and run training and evaluation — with no Axon individual viewing transcript text or listening to customer audio. Axon personnel review only de-identified numbers and evaluation results from Axon's own evaluation materials. Raw customer evidence audio and customer transcript text used for training do not leave the customer environment; training clips are short-lived, inaccessible to individuals, and automatically deleted; and this use case does not modify original customer evidence or live transcription behavior.