

Batch transcription output schema

Top-level object

Field	Type	Meaning
jobName	string	TranscriptionJobName.
status	string	COMPLETED on files written after a successful job.
results	object	Transcript payload.

results

Field	Type	Meaning
transcripts	array	Usually one element. Combined plain text of both channels, not separated by speaker.
transcripts[].transcript	string	Full mashed text including punctuation.
channel_labels	object	Present only with ChannelIdentification. Items regrouped by channel.
items	array	Every word and punctuation mark, each tagged with channel_label. Has numeric id.
audio_segments	array	Utterance chunks. Prefer this for turn-taking.

results.channel_labels

Field	Type	Meaning
number_of_channels	number	2 for L/R stereo. Transcribe supports at most two channels.
channels	array	One object per channel, typically ch_0 then ch_1.
channels[].channel_label	string	ch_0 or ch_1.
channels[].items	array	Words/punctuation on that channel only. Same shape as results.items minus id.

Item — results.items[] and channel_labels.channels[].items[]

Field	Type	Meaning
id	number	On results.items only. Zero-based. Referenced by audio_segments[].items.
channel_label	string	ch_0 / ch_1.
type	string	pronunciation (word) or punctuation.
start_time	string	Seconds from media start, e.g. "0.46". Omitted on punctuation.
end_time	string	Seconds. Omitted on punctuation.
language_code	string	Optional.
alternatives	array	Candidate spellings. Index 0 is the chosen one. Extra entries only if ShowAlternatives was set.
alternatives[].content	string	Word or punctuation character.
alternatives[].confidence	string	0.0–1.0 as a string. Punctuation is typically "0.0".

results.audio_segments[]

Field	Type	Meaning
id	number	Segment index.
transcript	string	Plain text of this utterance.
start_time / end_time	string	Segment bounds in seconds.
channel_label	string	Channel that produced the segment.
items	array of int	IDs into results.items that make up the segment.

Example (trimmed production job)

Shape only — arrays shortened. Times and confidence are strings. Punctuation has no timestamps.

```
{
  "jobName": "01634A8C_65E4B73B_3B279808_3B279808",
  "status": "COMPLETED",
  "results": {
    "transcripts": [
      { "transcript": "Hi mate. Hi mate, you're right. Yeah, I hear you. ..." }
    ],
    "channel_labels": {
      "number_of_channels": 2,
      "channels": [
        {
          "channel_label": "ch_0",
          "items": [
            {
              "type": "pronunciation",
              "channel_label": "ch_0",
              "start_time": "0.46",
              "end_time": "0.62",
              "alternatives": [{ "confidence": "0.915", "content": "Hi" }]
            },
            {
              "type": "pronunciation",
              "channel_label": "ch_0",
              "start_time": "0.62",
              "end_time": "0.899",
              "alternatives": [{ "confidence": "0.607", "content": "mate" }]
            },
            {
              "type": "punctuation",
              "channel_label": "ch_0",
              "alternatives": [{ "confidence": "0.0", "content": "." }]
            }
          ]
        },
        {
          "channel_label": "ch_1",
          "items": [
            {
              "type": "pronunciation",
              "channel_label": "ch_1",
              "start_time": "1.299",
              "end_time": "1.36",
              "alternatives": [{ "confidence": "0.376", "content": "Hi" }]
            },
            {
              "type": "pronunciation",
              "channel_label": "ch_1",
              "start_time": "1.36",
              "end_time": "1.5",
              "alternatives": [{ "confidence": "0.8", "content": "mate" }]
            },
            {
              "type": "punctuation",
              "channel_label": "ch_1",
              "alternatives": [{ "confidence": "0.0", "content": "," }]
            }
          ]
        }
      ]
    }
  ]
},
{
  "items": [
    {
      "id": 0,
```

```

    "type": "pronunciation",
    "channel_label": "ch_0",
    "start_time": "0.46",
    "end_time": "0.62",
    "alternatives": [{ "confidence": "0.915", "content": "Hi" }]
  },
  {
    "id": 3,
    "type": "pronunciation",
    "channel_label": "ch_1",
    "start_time": "1.299",
    "end_time": "1.36",
    "alternatives": [{ "confidence": "0.376", "content": "Hi" }]
  }
],
"audio_segments": [
  {
    "id": 0,
    "transcript": "Hi mate.",
    "start_time": "0.289",
    "end_time": "1.049",
    "channel_label": "ch_0",
    "items": [0, 1, 2]
  },
  {
    "id": 1,
    "transcript": "Hi mate, you're right.",
    "start_time": "1.129",
    "end_time": "1.94",
    "channel_label": "ch_1",
    "items": [3, 4, 5, 6, 7, 8]
  }
]
}

```

Notes

- **L/R mapping.** ch_0 is the first channel of the media (left in standard stereo); ch_1 is the second (right). Confirm with a known L-only fixture if agent/caller assignment matters.
- **Use audio_segments or channel_labels for turns.** transcripts[0].transcript is a mash of both speakers.
- **Do not assume results.items is globally time-sorted.** When channels overlap, a later id can have an earlier start_time. Sort by time, or walk audio_segments / channel_labels.channels[].items.
- **Overlap is allowed.** Both speakers talking at once produces overlapping timestamps on the two channels.
- **Numeric values are strings.** Parse start_time, end_time and confidence as floats.
- **This is not Call Analytics.** No ParticipantRole, sentiment, categories, or issues. Speaker diarization (spk_0) is a different setting and is not combined with channel ID on the same job.
- **Optional request features** (redaction, vocabulary filter, language identification, alternatives) add extra keys. They are absent unless requested.