Transcribe

The `Transcribe` element in Call Studio can be used to engage the Google Speech-to-Text services. The `Transcribe` element is located under the **Customer Virtual Assistant** group in the **Call Studio Elements**. This element is extension of the `Form` element and it engages the Speech Server resource on VVB to communicate with the Google Speech-to-Text Server. To indicate the Speech-to-Text server resource requirement, Call Studio creates a specific grammar - **builtin:speech/transcribe** - and sends it to VVB in a VXML Page. It does not specify which transcribe service is to be used; this is configured in VVB.

Note
- After playing non barge-in prompt for 120 seconds, VVB barges and creates recognition session for the Transcribe element. This causes the prompt to pause and skip a few seconds if the non barge-in prompt is longer than 120 seconds.
- The `Transcribe` element works both with Cisco DTMF and Nuance ASR adaptors.

Settings

Name (Label)	Type	Req'd	Single Setting Value	Substitution Allowed	Default	Notes
Input Mode	string	Yes	true	false	voice	The type of entry allowed for input. Possible values are `voice` (only voice input) and `dtmf+voice` (voice and DTMF input).
NoInput Timeout	int ≥ 0	Yes	true	true	5s	The maximum time allowed for silence before a noinput

						event is thrown. Possible values are standard time designations including both a non-negative number and a time unit. For example, `3s` for seconds or `3000 ms` for milliseconds.
Max NoInput Count	int ≥ 0	Yes	true	true	3	The maximum number of noinput events allowed during input capture. Possible values are `int > 0` where 0 = infinite noinputs allowed.
Max NoMatch Count	int ≥ 0	Yes	true	true	3	The maximum number of `NoMatch` events allowed during DTMF input capture. Possible values are `int > 0` where 0 is infinite `NoMatch` events allowed.
DTMF Grammar	string	Yes	true	true	None	This option is mandatory only if the input mode selected is DTMF and voice. It supports cisco DTMF regex.
Secure Logging	boolean	Yes	true	true	false	Whether or not to enable logging of potentially sensitive data of the element. If set to `true`, the element's potentially sensitive data will not be logged.
Terminiation Character	string	No	true	true	#	Terminate the voice stream or DTMF collection.
Max Input Time	int ≥ 0	Yes	true	true	30s	The maximum time (in seconds) the voice input is allowed to last. Possible values are positive integer values followed by s. For example, **50s**. Default value is `30s`.
Final Silence	int > 0	Yes	true	true	1s	The interval of silence (in seconds or milliseconds) that indicates the end of speech. Possible values are positive integer values followed by either s or ms. For example, **3s** and **3000ms**. Default value is `1s`.

`Recognize.phraseHints`	String	No	`true`	`true`	None	This is comma separated string that lists the hints for recognition. Hints are used to recognize a phrase or a word that is pronounced differently. For example, Savings, Current.
`Recognize.alternateLanguages`	String	No	`true`	`true`	None	Comma separated string of up to 3 additional BCP-47 language tags, listing possible alternative languages of the supplied audio other than the default language. For example, en-US, en-IN.

Custom VoiceXML Properties

Name (Label)	Type	Notes
`Recognize.NBestCount`	Integer	Specifies the maximum number of SpeechRecognitionAlternative messages within each SpeechRecognitionResult. The server may return fewer than max_alternatives. Valid values are 0-30. A value of 0 or 1 returns a maximum of 1. If omitted, returns a maximum of 1.
`Recognize.regionId`	String	Specifies the region to be used by the cloud speech-to-text transcription. This property should be configured in the root document of the project.
`Recognize.singleUtterance`	Boolean	Indicates whether this request should automatically end after speech is no longer detected. If this parameter is enabled, cloud speech-to-text will detect pauses, silence, or non-speech audio to determine when to end recognition. If this parameter is disabled, the stream will continue to listen and process audio until either the stream is closed directly, or the stream's length limit is reached. The default setting for this parameter is `true`.
`Recognize.model`	String	This is used to specify the machine learning model to be used by the cloud speech-to-text transcription to improve the recognition results. For example, see https://cloud.google.com/speech-to-text/docs/basics
`Recognize.modelEnhanced`	Boolean	Indicates whether enhanced model has been enabled. If it is enabled, the cloud speech-to-text transcription uses an enhanced

	speech recognition model to recognize speech and produce audio transcription more accurately.
	The default setting for this parameter is `true`.
	You can enable or disable data logging for enhanced speech model. For more information on data logging for enhanced speech model, see https://cloud.google.com/speech-to-text/docs/enhanced-models .

Element Data

Element Data	Notes
`value`	Transcribed text or DTMF collected.
`input_type`	Indicates the type of input captured (`dtmf` or `dtmf+voice`).
`confidence`	The speech recognition confidence between 0.0 and 1.0. A higher number indicates a greater probability that the recognized words are correct. The default of 0.0 is a sentinel value indicating that confidence was not set.
`language_code`	The language code that was triggered during recognition. Also see `Recognize.alternateLanguages` under Settings

Exit States

Name	Notes
`done`	Transcription completed.
`max_noinput`	The maximum number of `NoInput` events has occurred. If this is `0`, this exit state will not occur.
`max_nomatch`	The maximum number of `NoMatch` events that has occurred. This exit state will not occur if the maximum number of `nomatch` events is `0`. and `input_type` is `voice`. If `input_type` is `dtmf`, `max_nomatch` is the maximum number of DTMF mismatch with DTMF Grammar regex.

Audio Group

Form Data Capture

Name (Label)	Required	Max1	Notes
`initial_audio_group` (Initial)	Yes	Yes	Played when the voice element begins.
`nomatch_audio_group` (NoMatch)	No	No	Played when a `NoMatch` event occurs. This is applicable only when the input type selected is DTMF and voice.
`noinput_audio_group` (NoInput)	No	No	Played when a `NoInput` event occurs.

End

Name (Label)	Required	Max1	Notes
`done_audio_group` (Done)	No	Yes	Played when the form data capture is completed and the voice element exits with the `Done` exit state.

Folder and Class Information

Studio Element Folder Name	Class Name
Form	`com.audium.server.voiceElement.form.`

Events

Name (Label)	Class Name
Event Type	You can select **Java Exception**, **VXML Event**, or **Hotlink** as event handler for this element.