

Speaker Separation

Automatically Identify Customer & Agent Speakers Within Single Channel Audio Recordings

CallMiner Speaker Separation is a voice biometrics-based software that divides mono recordings into speaker channels representing the agent and the customer portions of a call to improve speech analytics effectiveness.

Identify Speakers in Mono Recorded Contact Centres

Many contact centres do not record calls in stereo due to limitations of 3rd party recording equipment, analogue trunking or lack of VOIP support. Or, it may simply be due to monaural compression applied to conserve storage or other factors. As a result, a single channel is used for audio from the speaker and the agent. Mono recording poses challenges for speech analytics. The primary issue is that transcriptions cannot identify the source of speech between the agent and caller. As a result, it becomes difficult to zero in on customer satisfaction or agent performance. Mono recording makes it impossible to pinpoint whether the caller or agent is responsible for what was said as well as the associated sentiment and acoustic measures.

CallMiner Speaker Separation overcomes this challenge with voice biometrics. Voice characteristics are used to “voiceprint” an agent. CallMiner Eureka can then clearly identify caller and agent speakers within its powerful analytics engine. As a result, automated agent scoring, caller and agent sentiment metrics, topic discovery based on what callers or agents are saying and more is enabled.

How Speaker Separation Works

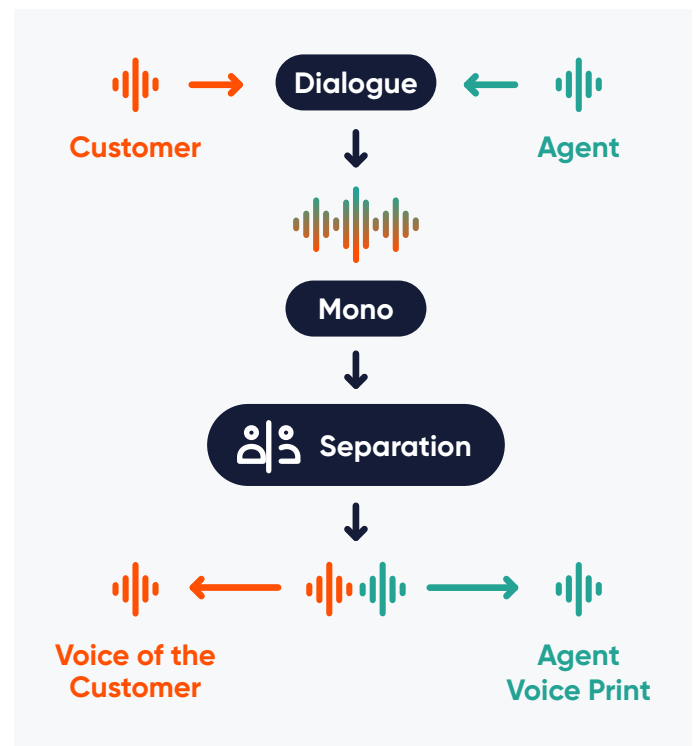
A “passive enrollment” process is used by training on a group of calls with the same agent. The system then identifies the most prevalent talker across those calls (assumes the customer changes from call to call) and assigns the agent’s voiceprint. During the speech-to-text process, each part of the conversation is then attributed to the correct speaker – improving transcription readability, reducing the time to target agent or customer issues via speaker filtering, and speaker targeted topic-mining.

Voiceprinting Agents

CallMiner Speaker Separation is trained to identify which speaker is the agent with a minimum of 10 recordings ideally of at least 1 minute of audio or more for each call. Voiceprints are passively captured, which means that calls in the normal course of dialogues are biometrically monitored without agent or customer impact.

Biometric Assignment

The speaker separation algorithm makes two passes through an audio recording. First, speakers are divided into A & B “channels” utilising a separation process. A voiceprint is then captured for the agent based upon biometrics across multiple calls. The other speaker is assigned as the customer.



Once a voiceprint is created, all subsequent audio for that agent is processed through speaker separation identification. Thereafter, the system assumes any speaker NOT the agent is the customer.

Voiceprinting Agents

Transcript Usability

Speaker-associated search with tags automatically applied for categories are enabled with speaker separation. Issues such as customer satisfaction and escalations are easily identified. Agent performance including compliance and sentiment is also clear with automated scoring

Topic Discovery

Trending issues that may or may not have been identified are revealed based on agent or customer utterances. Topic circles with size indicating call volume split between agent and customer speakers can innovatively support root cause awareness and action

Accuracy

Speaker Separation between and agent and a caller is highly reliable if good quality audio is available along with the following considerations:

- Call duration longer than 5 seconds
- 3rd person talking or background over talk – assigns voices to most similar agent or customer
- Hold music, especially with a heavy percentage of vocals – use instrumentals

Efficiency

Storage requirements are the same for mono as only the transcripts are separated between caller and agent. Voiceprint processing overhead is likely 5% or less, compared to 25% and perhaps significantly more required for stereo call recording. Also eliminates the need for a stereo call recoding upgrade.

Unobtrusive

Passive voiceprint enrollment means agents always remain in service. Also, the integrity of transcription content remains only not associated with identity if speaker separation fails.



New Agent Enrollment Automation

CallMiner Speaker Separation automatically begins its passive voiceprint enrollment when a new agent metadata notification is received. The system will begin to compile the proper number of calls required to complete voice printing for that agent. Until that process is complete, new agent calls will be captured without speaker identity.

Request a demo today

For more details about Speaker Separation, contact your CallMiner Sales Director or:

sales@callminer.com