Building a Speech Solution for the Contact Center:

Key Components & Considerations

Speech analytics solutions powered by speech-to-text (STT) technology can allow contact centers to unlock insights that produce outcomes like increased customer satisfaction and revenue. By using insights from post-call or real-time transcription, contact centers can:

— Reduce customer churn

— Identify targeted (real-time) upsell opportunities

— Ensure agent compliance

— Identify and enable (real-time) agent training and coaching opportunities

— Reduce average handling time

— Discover unexpected product and service strengths or weaknesses

However, designing and deploying a speech analytics solution for a contact center isn't as straightforward as you might expect, with many components and considerations to account for.

Critical Considerations for Contact Centers

As you plan or build your contact center speech solution, four aspects should be considered with every technology decision.

1. Scalability:

The end goal is a solution that can be scaled easily and affordably as the business and product use grows. If you choose an on-premises solution, consider how to minimize the hardware that will be required as you scale.

2. Flexibility

Real-time transcription allows for the most potent insights and game-changing behaviors in contact centers. Even if post-call transcription is the initial focus, for maximum flexibility, contact centers should also be ready for real-time STT. The more flexible a solution is from the start, the less need to restructure technology down the road as a contact center's tech stack and business needs evolve.

3. Security & Compliance:

Audio is a valuable asset that can contain sensitive information. An important consideration is whether you are comfortable sending audio to the cloud or if you prefer an on-premises solution. Either way, you want to know how the technology you use protects your data. Other common security and compliance concerns include audit trails and PCI redaction from audio and transcripts.

4. Total Cost of Ownership:

For cloud-based solutions, usage primarily determines pricing. For on-prem solutions, hardware costs – and how growth will impact those costs – must also be factored in. (The higher throughput you have per server, the lower your cost.) Time to value is another aspect to be weighed with product quality, flexibility, and scalability.

Essential Components of a Speech Solution
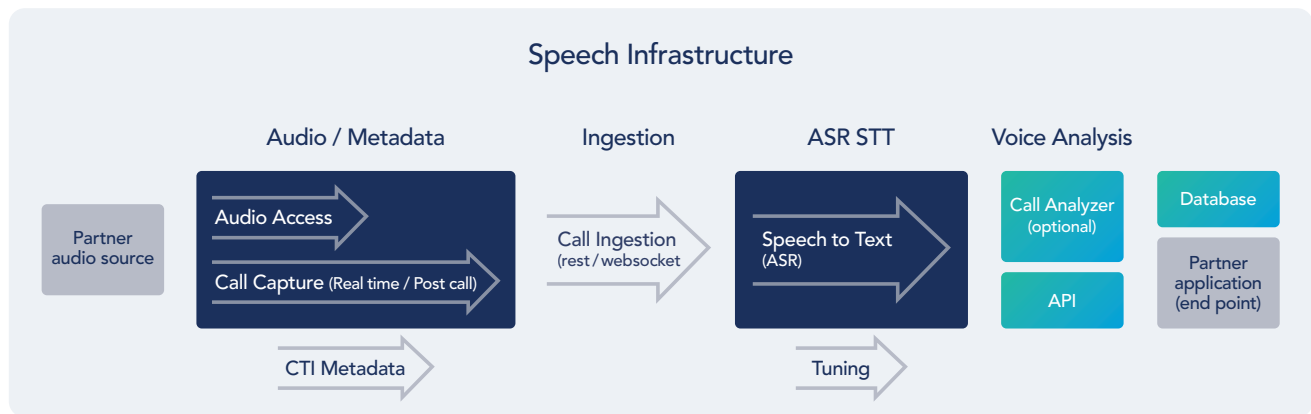
Many teams planning to introduce a speech solution start by looking for automated speech recognition (ASR) and STT technology. While this is a vital piece of the puzzle, as you see above, it is far from the only component of a complete speech solution.

> **Audio quality tip**
>
> Degraded audio means degraded insights. To protect your audio quality, always record in channel-separated stereo and minimize conversions across the audio lifecycle.
>
> **More about optimizing audio quality**

Audio: Call Access & Call Capture

Accessing call audio can be tricky, and how you do so is tied to your method for call capture. Your primary choices are to get call recordings out of a call recorder or tap directly into a telephony system.

While getting recordings from a call recorder involves less development work, a key benefit of tapping into a telephony system is the opportunity to deliver real-time guidance and insights to agents and managers as calls occur. DTT, or direct-to-transcript, solutions can also simplify the entire diagram above. With a DTT solution, you can streamline call access, call capture, call ingestion, and speech-to-text into one united process and then send the transcripts straight to your analytics application.

Method	✅ Pro	➖ Con
Getting recordings from call recorder	✅ Faster setup with less internal development work	➖ Call recorder companies charge expensive fees for audio access
		➖ Audio often low quality, limiting the quality of analytics / insights derived
		➖ Real-time transcription or insights impossible
Tapping into a telephony system	✅ Better audio allows for more advanced analytics	➖ Higher effort deployment due to telephony system complexity
	✅ No roadblocks or large fees from recording vendors	
	✅ Allows for real-time, direct-to-transcript (DTT) solutions	

Orchestration: Call Ingestion

Call orchestration and ingestion is the process of how you queue and ingest call audio into your ASR engine. The exact way in which you do this will be specific to your system; for post-call transcription, custom code will be required to send audio to the ASR engine via an API.

For post-call transcription, one common method is to write audio files to a directory and create software that will upload any new files added to that directory to the ASR engine. Alternatively, you may choose to connect to the call recording system API to extract and send the call audio to the ASR engine.

Call ingestion tip

Need enhanced security? Keep call ingestion on-prem by setting up a direct-to-transcript (DTT) system with servers inside the data center that houses your telephone system.

More about keeping your call data secure

Automated Speech Recognition: Speech-to-Text & Tuning

Automated speech recognition (ASR) technology for contact centers has become more affordable, powerful, and reliable in recent years. Best-in-class products combine fast transcription speeds with robust feature offerings and flexible deployment and integration options. For real-time applications and in-call analytics, low latency is essential. High throughput per server is also important for lowering cost.

For ASR product completeness, look for STT products offering real-time and post-call transcription with punctuation, capitalization, number formatting, security features, multiple languages supported, speech metadata (emotion, gender, etc.), ingestion of Computer Telephony Integration (CTI) metadata and custom tuning services. Tuning an ASR engine will increase word recognition and improve accuracy of high-value product, brand, and industry terms. This type of accuracy is known as specific word accuracy and is actually a more critical measurement for business analytics tools than overall word accuracy. Tuning is also useful for better assessing call and agent sentiment and adjusting what sensitive information is redacted for compliance and security.

ASR engine selection tip

When comparing ASR engines, look at a vendor's track record of improving STT accuracy, lowering costs, adding new language models, and enhancing ease of use over time.

More about choosing an ASR engine

Tuning tip

Find an ASR engine with built-in language models for your industries and/or applications to reduce time and added cost spent on custom tuning and get better STT accuracy and analytics faster.

More about improving STT accuracy

Call Metadata

Contact centers introducing a voice-of-the-customer speech analytics tool into their organization will need to utilize call metadata such as agent name, agent team, and caller number. This metadata is distinct from any metadata that an STT tool may provide, which can include information on gender, sentiment, and emotion, for example.

To link this call metadata with call transcripts, contact centers must first investigate how to extract call metadata from the CTI system. Once you understand that process, there are two primary choices: a pre-built solution which allows you to get to market quicker, or the route of building your own. Your ASR engine provider or analytics software may offer a solution that meets your needs and works with your CTI system. If not, you will have to build a custom ETL system, using an internal database to relate each call recording or transcript with its appropriate call metadata.

Call metadata usage tip

An analytics system with access to call metadata and call transcripts combined can deliver enriched insights without the complexity of customization and telephony integration skillsets.

More about the new age of contact centers

Voice Analysis

A call analyzer is an optional analysis tool provided by some ASR engines that allows organizations to easily organize, filter, search, classify, visualize, and report on call data. A call analyzer can allow you to search transcripts for specific words or phrases as well as automatically tag calls based on designated criteria. In particular, call analyzers are useful for descriptive analytics, allowing users to view statistics, trends and changes, in metrics such as call volume, duration, agent emotions, and client emotions over the short- and long-term. Other information such as the agents with the most and least silence, talk, and overtalk time may also be available for review.

Whether or not you have a call analyzer as part of your ASR system, you will need to connect the ASR engine to your analytics or business intelligence application via an API. This process should be straightforward if you've selected an ASR engine with open architecture that is optimized for ease of integration.

How to Get Started on Developing Your Speech Solution

Voci Technologies offers a best-in-class accurate, flexible, and enterprise-scale ASR engine for real-time or post-call transcription and is a trusted partner to industry leaders such as **Sykes Enterprises**. Voci's technology experts have over a decade of experience developing AI-powered STT solutions that have serviced thousands of customers and transcribed millions of hours of audio.

The End Product:
Bringing it Together to Deliver Actionable Insights

While a call analyzer provides useful information, leading contact centers go beyond creating dashboards and reviewing metrics. The greatest opportunity to differentiate your organization comes from using STT solutions for predictive and prescriptive analytics. These more advanced analytics allow organizations to first predict customer needs and responses and then to recommend next best actions for agents and managers in real-time. By introducing a speech solution with predictive and prescriptive analytics, you can drive better customer experiences, smarter business decisions, and significant cost savings for the contact center.

Not only can Voci's ASR engine offer the lowest total operating cost available in the market, the company's solutions are designed to simplify your entire speech infrastructure – and experts who can advise you on every component of your speech solution.

**Contact us today at
412-621-9310 or info@vocitec.com
to see how we could help you.**

Voci Technologies, the leading speech analytics platform provider, enables enterprises to gain actionable insights on their terms from 100% of customer calls. For information, visit www.vocitec.com.

Real-Time Speech-To-Text at Enterprise Scale

vocitec.com