



Use Cases **Differ Between Text and Voice**

While chatbot vendors are now trying to offer an embedded solution that contains text and voice, these models cannot be clubbed into one platform.

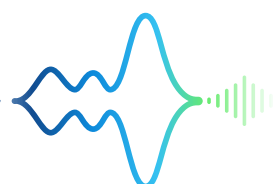
By 2022, close to 200 million jobs would be lost globally due to the Covid crisis. A lot of those unemployed will need to make changes in their monthly payments like home loan tenure, convert their credit bills into EMIs, and remove value-added services. Such customers connecting with a company during an emotionally volatile state may not just be looking for a solution, but could also be seeking a sympathetic ear.



In such a situation, a vulnerable customer will prefer speaking to someone to differ payments rather than type a series of requests for each liability. For instance, converting a credit bill into an EMI is one command that is executed after typing out a few details. Then comes home loan tenure increase, which requires another set of instructions. Here, it is faster and smoother for the information to be captured via voice.

Let's take another example. A customer who lost a parent to Covid can file a death claim online. But speaking to a service executive who could empathetically listen to their concerns could soothe nerves during distress. Not only can the voice-led channel help minimise claim delays by specifying the exact documents needed, but the customer can also understand the formalities over a single call.

We have read how the current customer service models are missing out on the primacy of voice. There is a perception in the market that having a single solution for text and voice will help bridge the gap. But simply building a voice solution over existing text solutions may hamper the user experience. In customer service, voice is designed to understand the nuance and gravity of a request. This is true especially for emergency situations where customers may not have the time nor the mind space to sit and type requests like finding a network hospital or an unauthorised transaction through a bank account.



Trivializing voice and offering it as a 'good-to-have' solution by chat providers is counter-intuitive because voice is a specialized solution that encompasses the layers chat requires, plus catches peculiar behaviours like tone and pauses in speech. The demand for Voice AI has grown exponentially in the past few years.



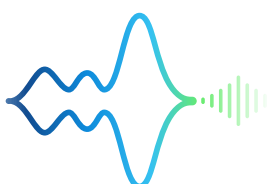
As per a study by Opus Research, 84% of businesses in eight key industries are enabling voice mobile apps.



67% of top brands are implementing voice assistants. And customers, too, have grown accustomed to voice assistants.

According to a report by Statista, the number of digital voice assistants is likely to reach 8.4 billion units by 2024. So, it makes sense that companies want to adapt to this growing trend.

Voice is convenient, especially because humans speak and perceive things differently over speech than text. For instance, an indecisive food-delivery customer who keeps changing his/her order may find it easier to finalise an order over voice rather than typing and selecting products. Having a voice conversation also enables them to make a faster decision on what food to order.



A 2018 report by Gartner said that organizations have spent almost a decade integrating channels to provide users with seamless access, but that omnichannel approach has left 85% of organizations with fragmented customer engagement channels, leading to inconsistent customer experience.

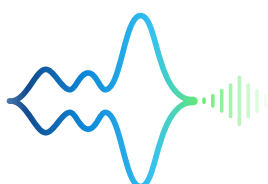
While the thrust is on 'omnichannel' presence by brands, deploying voice effectively could help resolve a lot of customer complaints across product and service categories. Being present across customer touchpoints is good, but resolving queries constructively and consistently on a single platform — voice — is better.



How is voice different from text?

Chabots follow a flow wherein the text input is fed into the spoken language understanding engine. This engine understands the input/query and decides on the next course of action. Based on the context of the conversation, the response is prepared in a text format but relayed back to the user.

Voice AI, on the other hand, has two engines specifically available to understand speech. One is a speech-to-text engine, and the other is an automatic speech recognition engine.



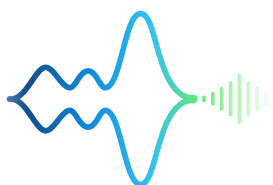
SLU - The Spoken Language Understanding

NLU - Natural Language Understanding engine

When a user speaks, their voice is converted into text through the speech-to-text or STT Engine. Next comes the **Spoken Language Understanding (SLU)** or **Natural Language Understanding (NLU)** engine, which can be defined as the brain of the Voice AI systems. This engine extracts the data from the text produced by STT and decides on the next course of action. This action could be a voice reply, text message for authentication purposes, or even sending pre-requested documents to customers.

The last part of this process is the dialogue manager, which acts as the orchestrator of the entire conversation. This is the block that manages the flow of data among the above three blocks and the flow of the conversation. And all these processes happen within milliseconds over the cloud, so it is device agnostic.

The end goals of voice and text are also fundamentally different. Text is intended to resolve basic customer requests and redirect complicated questions to customer service personnel. For instance, a customer looking to book a restaurant table is able to ask multiple questions in one go through voice. These could be the waiting times at certain points of the day, the chef's menu, and specific details about the dishes (ingredients, spicy, vegan alternatives, etc.). An added layer of benefit are newer concepts like paralinguistics being used in the Voice AI ecosystem. This involves communication other than spoken words, including tone, pitch, pauses, and gestures. For sales teams of customer-facing brands, this offers a tremendous opportunity to gauge a customer's interest in the product and gauge their buy intent.



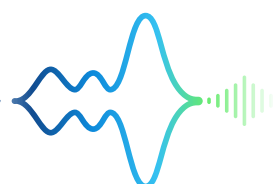
Once Voice AI determines who is more inclined to buy a product/service, additional time can be spent to explain to convince the customer.

This essentially means that cross-selling products will be far easier and effective if these Voice AI solutions are deployed. Some sectors that could take advantage of this concept are hospitality chains, restaurants, and financial institutions selling retail products like credit cards and quick personal loans.



It is often noticed that customers need to be nudged to reveal information, a process that can be done effortlessly over voice. Say, a newly launched shoe brand wants deeper feedback on the products. Using the customer database, a caller could be contacted using Voice AI to seek a detailed response on the pros and cons of the shoes. A customer may like the product quality but may have found its pricing to be steep while another customer may be looking for newer colour options. Customers seldom fill long review forms that are sent post-purchase, hence bringing voice into this equation helps in better assessment.

Based on the collective feedback, companies will also be able to tweak their product offerings accordingly, leading to improvement in sales. Customers, too, feel satisfied that their opinions have been taken into consideration.





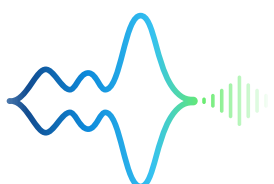
A clubbed solution isn't effective

Voice is an ideal turf for AI to learn, evolve, and constantly upskill by taking due note of user sentiments and emotions. And the best part? The user doesn't need to be able to write a language fluently. Voice AI provides the unmatched ability to interact through casual conversations.

Critical user feedback, including anger, can't be spotted immediately on text. This is essential for companies involved in product development, where continuous feedback generation is the key to success. As stated earlier, chatbots rely on key terms such as bad, poor, or terrible to deduce that the experience is unsatisfactory. Voice, on the other hand, listens attentively to different users to understand their sentiments.

In industries such as telecom, for example, users often use sarcasm to point out service delays. "Wow! You guys are amazing. It only took you six months to resolve simple network issues" will be interpreted as 'good feedback' by a chatbot because the sentence uses terms like 'wow' and 'amazing'. But a Voice AI will be immediately able to spot the tone and offer better resolution.

Vendors offering 'text+voice' combo products do not understand the performance requirements of Voice AI systems. Low latency or quick processing of data to offer the right answers is crucial. Right now, there seems to be a rush among brands to implement AI for customer service. But the key here is to operationalize a solution that is accurate and solves a given problem. The thing to remember is that context and slang change with geographies. They are different in different markets. This means each Voice AI system needs to be modified to suit the audiences in that location. This is where the expertise of market providers, such as Skit, comes in handy.



220+ - people in our team

70+ - machine learning engineers

Not only can Skit help provide specific voice solutions for enterprise customers, but our agile systems can fine-tune each process for brands. With **our team of 220+ people**, including **70+ machine learning engineers**, we know what works in each sector.

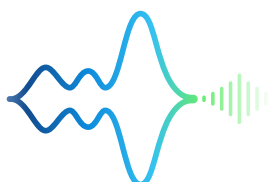


The emerging dynamics of voice

Voice works best for context-led conversations where tone and inflection can convey a response without using actual words. And as the technology develops, its use-cases have also been evolving. In areas like sales and product testing, Voice AI could be used to pitch the product better and sound more persuasive.

Customers are also more likely to interact for a longer duration with a Voice AI system that understands his/her specific needs. These conversations are also useful for training the internal systems and for conducting quality checks at a later stage. For example, a fintech company developing a buy-now-pay-later (BNPL) product could use an advanced Voice AI system to capture the purchasing patterns of a customer. Since it is responsive, the customer can also cross-question Voice AI on the relevance of the terms and conditions of the BNPL feature and default penalties.

And if the Voice AI notices that target customers are enquiring repeatedly about penalties, this can be relayed back to the brand so that its messaging can be tweaked to include the terms upfront.



In markets like the US where financial frauds lead to brands losing millions of dollars in revenue and also reputation loss, Voice AI could come handy in adding a layer of voice-led authentication.

\$0.58 lose

A report by Pindrop Labs said that the average call center loses \$0.58 to fraud for every incoming call.

\$14B lose

That means that companies across the US lose \$14 billion to fraud attacks annually.

Here, deploying voice to recognize and identify customer details will help prevent such risks. This is because the AI can identify regular pauses and also spot any nervous tones indicating the presence of fraudsters on the call.

Psychological concepts like entrainment could be complementary to the existing services where customer interactions can be improved. For instance, an angry customer could be pacified through Voice AI speaking in a calmer voice tone. Similarly, if a customer will be understood even if he/she switches to a different language midway into the call.

Voice solutions are getting richer. While a lot of vendor solutions already exist in the market, specialized products are far and few in between. A Voice AI product that is constantly tested for different use-cases across sectors is what will be suitable for commercial use.

