

Speech TECHNOLOGY

Speech Recognition Customer Self Service Virtual Assistants Analytics Artificial Intelligence

More Topics ▼ Industry Solutions ▼

HOME SUBSCRIBE ▼ NEWS IN DEPTH ▼ WHITE PAPERS INDUSTRY RESEARCH WEBINARS RESOURCES ▼

CONFERENCES ▼ ABOUT ▼ 



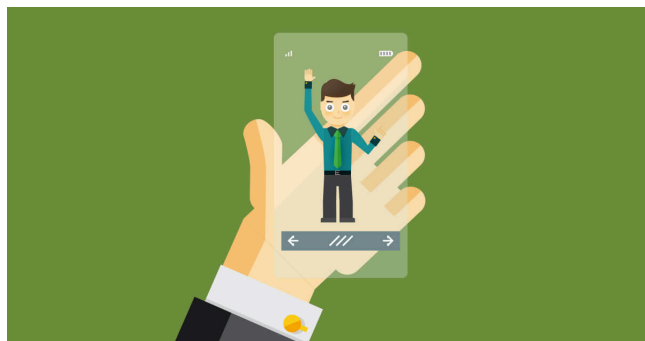
December 3, 2024

By [James A. Larson](#) program co-chair, SpeechTEK 2021
Forward Thinking



Site Sponsors

Shopping Made Easy: Features of a Modern Conversational Assistant



Gone will be the days of navigating complex menus and endless prompts. Enter the conversational assistant. your personalized

WEB EVENTS

Avatar Platforms in Customer Service
Coming March 11, 2025

Ethical Implications of Voice Generation
Coming July 15, 2025

Driving Speech Technology Trends with AI
Coming October 14, 2025

[More Web Events](#)

POPULAR ARTICLES

GET SPEECHTECH EWEEKLY IN YOUR INBOX - SIGN UP FOR FREE

Enter Email Address

GO

Below is an imaginary interaction between me and a conversational assistant named Anna, with the key features that will come to define a modern conversational assistant shown in brackets:

I use my Android smartphone [1. **Device independence**, to seamlessly connect with Anna using a smartphone, smart speaker, or any other device] to connect to the Anna conversation assistant [2. Using a **directory of available conversational assistants**, which locates and connects with the relevant assistant].

I see text and hear an animated avatar that says, "Welcome to Anna's Fresh Veggie Farm Stall Conversational Agent. How may I help you?" This verifies that I have reached the desired conversational assistant.

The dialogue pauses, waiting for me to respond. "Good morning. I plan to make a tomato salad for dinner. Can I get four tomatoes?" I ask.

Anna hears my spoken request and determines that I want to speak and listen rather than type and read [3. **Multimodal communication** uses voice, text, video, and/or images to converse with the conversational agent].

Anna determines that I speak English [4. **Language identification** that determines the customer's natural language], so she responds using my preferred language. If you speak your requests, Anna will respond clearly, creating a natural conversation flow [5. **Voice recognition, voice synthesis, and natural language processing**].

Anna next determines who I am [6. **Speaker identification** determines who is speaking], verifies my identity using both fingerprint scanning and facial recognition [7. **Multifactor authentication**, to verify identity], and then addresses me by my name [8. **Secure user**

SoundHound and Lucid Motors Partner on In-Vehicle Voice Assistant with GenAI

AI Language Translator Tool Market to Soar with 26.2 Percent CAGR

iWallet Launches iWallet Voice AI

Timekettle Releases W4 Pro Earbuds with Babel OS for Real-Time Two-Way Call Translation

GET SPEECHTECH EWEEKLY IN YOUR INBOX - SIGN UP FOR FREE

Anna says, "Good morning, James Larson, I recommend you start with these beautiful, ripe Roma tomatoes. They will be the star of the show." I see video of Anna as she initially selects four tomatoes from the tomato bin and sets them on the counter.

I grimace and groan when I notice that one of the tomatoes has a spoiled spot. Anna notices my reaction. [9. **Emotion detection** shows that Anna is sensitive to my reactions. If she senses hesitation, she'll be there to help, such as noticing your dislike for the bruised tomato.]

Anna tosses the offending tomato into the garbage bin behind her.

"Four more tomatoes and two shallots to add a bit of crunch," I continue.

Anna selects four more tomatoes and two fresh shallots and adds them to the pile on the counter.

"Does the salad need carrots?" I ask Anna.

"No, your son has an allergy to carrots. You should not use carrots!" Anna reminds me after examining my personal profile [10. **External knowledge**, meaning that Anna has a vast knowledge base that remembers your preferences and allergies (i.e., your son's carrot allergy) to personalize the experience].

After examining a book of recipes, Anna says: "Cucumber, bell pepper, avocado, or zucchini can add texture and variety to your salad. Fresh basil is a classic pairing with tomatoes, but you can also use oregano, mint, chives, or dill."

I respond, "I'll take two avocados and some chives."

As Anna assembles my order, she notes that feta or mozzarella cheese can add protein. "Let me check their availability with Paolo," Anna pauses while she privately connects to another

GET SPEECHTECH EWEEKLY IN YOUR INBOX - SIGN UP FOR FREE

the user is unable to see or participate in this conversation.]

Anna summarizes her discussion with Paolo:
“The Paolo conversational agent has both feta and mozzarella cheeses available. He assures me they are fresh. Is there anything else you need?” she asked.

“No thanks, that’s all that I need today. Please connect me to Paolo after we finish,” I say.

“May I deduct the amount due, \$14.56, from your credit card with the number ending in 8799 [12. **Secure payment**]?” I pay for my groceries conveniently and confidently with encryption and secure payment processes.

“Of course,” I agree.

“Done,” says Anna. “A receipt has been emailed to you.”

“Thanks a million,” I say.

Anna asks: “What should I do with the recording of this conversation? May it be used for training other conversational agents, may it be mined for business data, or should it be deleted?”

“Make it disappear,” I instruct.

Anna asks, “Do you now want me to connect you directly to the Paolo agent, who can recommend which cheese to add to your salad?”

“Sure. Thank you,” I reply.

Anna says, “Thank you for your business. You will now be connected to the Paolo conversational assistant, who is ready and excited to talk about cheeses.” [13. **Delegation** switches the conversation between conversational agents.] Anna transfers the connection to Paolo.

*GET SPEECHTECH EWEEKLY IN YOUR INBOX - **SIGN UP FOR FREE***

agent. (For additional information about delegation and mediation, see <https://voiceinteroperability.ai/>.)

Lastly, Anna and Paolo implement behind-the-scenes measures such as encryption, access controls, regular updates, and strong authentication protocols [14. **Security**, which protects the agent's data and functionality against unauthorized access, attacks, and malicious intent].

Will your automated agents be able to perform these feats in 2025? Only time will tell, but with the rapid advancements in AI, the future of shopping looks bright—and delightfully conversational.

James A. Larson, Ph.D., is an independent voice technology expert and can be reached at jim42@larson-tech.com.

It's Time for Speech Recognition to Get Some, Well, Recognition

The 2024 Nobel Prize in Physics was awarded to John J. Hopfield and Geoffrey E. Hinton for their pioneering work on artificial neural networks. Their research has significantly advanced the fields of machine learning and artificial intelligence by using statistical physics concepts to develop neural networks that can recognize patterns in large datasets. This technology is proved to be a cornerstone of modern AI.

It's about time that such work was recognized by the Nobel committee. It would be nice if significant advances in speech technology earned major recognition. May I suggest a new Speech Technology Hall of Fame award that has the prestige of a Nobel prize (if not the associated prize money)?

Below is my list of potential candidates:

GET SPEECHTECH EWEEKLY IN YOUR INBOX - SIGN UP FOR FREE

automatic speech recognition. Audrey could recognize the first 10 numbers (0-9) with an accuracy rate of about 90 percent, provided the speaker was male and spoke with a specific delay between words

The **Harpy** speech recognition system (1976), capable of understanding more than 1,000 words, similar to a 3-year-old's vocabulary, was developed by **Bruce Lowerre** and **Raj Reddy**.

Hidden Markov Models (HMMs) were developed independently by several researchers in the late 1960s and early 1970s. Some of the key contributors include **Leonard E. Baum**, a mathematician who developed the Baum-Welch algorithm, a fundamental algorithm for training HMMs; **Andrew Viterbi**, a computer scientist who developed the Viterbi algorithm, a dynamic programming algorithm used for decoding HMMs; and **Erkki Oja**, a Finnish computer scientist who made significant contributions to the theory and application of HMMs. HMMs form the bases of modern speech recognition.

SPHINX-I (1987), the first speaker-independent continuous speech recognition system to not require user training beforehand, was developed by **Kai-Fu Lee**, a protégé of AI pioneer **Raj Reddy**, at Carnegie Mellon.

Dragon NaturallySpeaking (1997), the first commercial speech recognition software, which popularized the use of voice input for dictation and other tasks, was developed by **Dr. James K. Baker** and **Dr. Janet M. Baker**. The software can transcribe spoken words into text with high accuracy, enabling users to dictate documents, send emails, browse the web, and perform other tasks without typing. Users are required to train the system before using.

Siri (2011), the first virtual assistant to use natural language processing to understand and respond to voice commands, was developed by **Tom Gruber**, **Adam Cheyer**, and **Dag Kittlaus**.

GET SPEECHTECH EWEEKLY IN YOUR INBOX - SIGN UP FOR FREE

devices, making calls, and providing information. It was developed by **Sundar Pichai**, **Jeff Dean**, and **John Giannandrea**.

ChatGPT (2022), the first widely available genAI chatbot, which was based, upon release, on the large language model (LLM) GPT-3.5, was developed by **Sam Atman** at OpenAI. It's designed to interact with users in a conversational way, providing information and completing tasks as instructed. It's trained on a massive amount of text data, allowing it to generate humanlike text in response to a wide range of prompts and questions.

Who would you induct into the Speech Technology Hall of Fame?

—J.L.



FREE

FOR QUALIFIED
SUBSCRIBERS

SUBSCRIBE

NOW 

CURRENT PAST
ISSUE ISSUES

GET SPEECHTECH EWEEKLY IN YOUR INBOX - **SIGN UP FOR FREE**

ALSO ON SPEECH TECHNOLOGY MAGAZINE

2 years ago · 1 comment

Microsoft Releases Tools That Will Enable

a year ago · 1 c

2024 Sta in the Sp Technolc

0 Comments

 1 Login ▼

Start the discussion...

LOG IN WITH

OR SIGN UP WITH DISQUS 

Name

 Share

Best Newest Oldest

RELATED ARTICLES

GET SPEECHTECH EWEEKLY IN YOUR INBOX - SIGN UP FOR FREE

Searchable Pitfalls Generative
 CAs Facing AI Is
 Will Conversational
 Make Assistants
 Conversational Hallucinations
 AI and Knife
 More Deepfakes for
 Convenient Today's
 As conversational assistants proliferate, they'll become harder to discover. Bad data (and actors) can have consequences. Helpful assistants can become even more so thanks to genAI.

09 MAY 2024

28 OCT 2024

19 MAR 2024

Speech TECHNOLOGY

SpeechTechMag.com is the premier online destination for comprehensive, independent coverage of information impacting speech technologies.

All Content Copyright © 2009 - 2024 Information Today Inc. 



Speech Technology Magazine
 143 Old Marlton Pike

Resources

Home
Speech Technology Magazine
 Digital Editions (PDF Download)
 Industry Research
 Speech Technology Newsletters
 Speech Technology Topic Centers
 Speech Technology Industry Solutions
 Speech Industry Awards
 Sponsored Videos
 Reference Guide
 About/Contacts

Other ITI Sites

Database Trends and Applications
 DestinationCRM
 Faulkner Information Services
 InfoToday.com
 KMWorld
 Enterprise AI World
 Online Searcher
 Smart Customer Service
 Streaming Media
 Streaming Media Europe
 Streaming Media Producer
 Unisphere Research

GET SPEECHTECH EWEEKLY IN YOUR INBOX - SIGN UP FOR FREE

*GET SPEECHTECH EWEKLY IN YOUR INBOX - **SIGN UP FOR FREE***