

CASE STUDY

Over 200k key phrases/prompts collected in 12 Global Languages

Shaip collaborates with a leading tech company to develop voice-activated systems for the automotive industry



COMPANY

A leading
Tech Company



INDUSTRY

Technology



USE CASE

Conversational AI:
Key Phrase Collection for
voice activated systems



OUR OFFERING

Key phrase/Prompts
Audio Collection

KEY STATS

No of Prompts
Collected
200k+

No. of
Languages
12

No of
Speakers
2800

Languages
Supported
**Norwegian, Danish,
Swedish, Czech,
Hebrew, UK English,
French, Italian,
Polish, Arabic, Turkish,
India - English**

Project
Timeline
**less than
35-37
Weeks**

OVERVIEW

The automotive industry has rapidly adopted voice-activated systems, with major players like Ford, Tesla, and BMW integrating advanced voice recognition in their vehicles. By 2022, it was estimated that over 50% of new cars featured voice recognition capabilities. These integrations aim to enhance safety, allowing drivers to operate navigation, entertainment, and communication functions without distractions. The market value for voice recognition in autos was projected to surpass \$1 billion by 2023, indicating a growing demand for hands-free, intelligent in-car interactions.



CHALLENGES

The client is a global leader in conversational intelligence who offers voice AI solutions that let businesses offer incredible conversational experiences to their customers. They were working with leading automotive companies to train their voice-activated systems with branded key phrases and needed Shaip's expertise in audio data collection. The critical requirements of the client were:



- » **Data Collection:** Collect over 200k prompts in 12 different languages in a stipulated time frame



- » **Crowd Sourcing:** Source 2800+ native speakers for each of the languages from different parts of the world



- » **Context & Intent Recognition:** To understand user requests correctly, systems needed to be trained on different variations for the same key phrase.



- » **Background Noise Handling:** Real-world scenarios involve background noises which may affect the output of ML model



- » **Audio format:** Ensure that the audio is recorded in 16khz 16bits PCM, mono, single channel, WAV file format with no audio processing



- » **Reducing Bias:** Biases in voice recognition can lead to dissatisfaction or even alienation. The client required voice samples from different demographics and geography w.r.t. age and gender that ensures inclusivity



- » **Quality Check:** All speech recordings will undergo quality assessment and validation, only validated speech recordings will be delivered. If Shaip does not meet the agreed Quality Standards, Shaip will redeliver data at no additional cost



- » **Recording Environment:** Recordings should have clean audio without background noise. Key Phrases to be recorded using normal speech. No variation in speech (e.g., fast, normal slow)

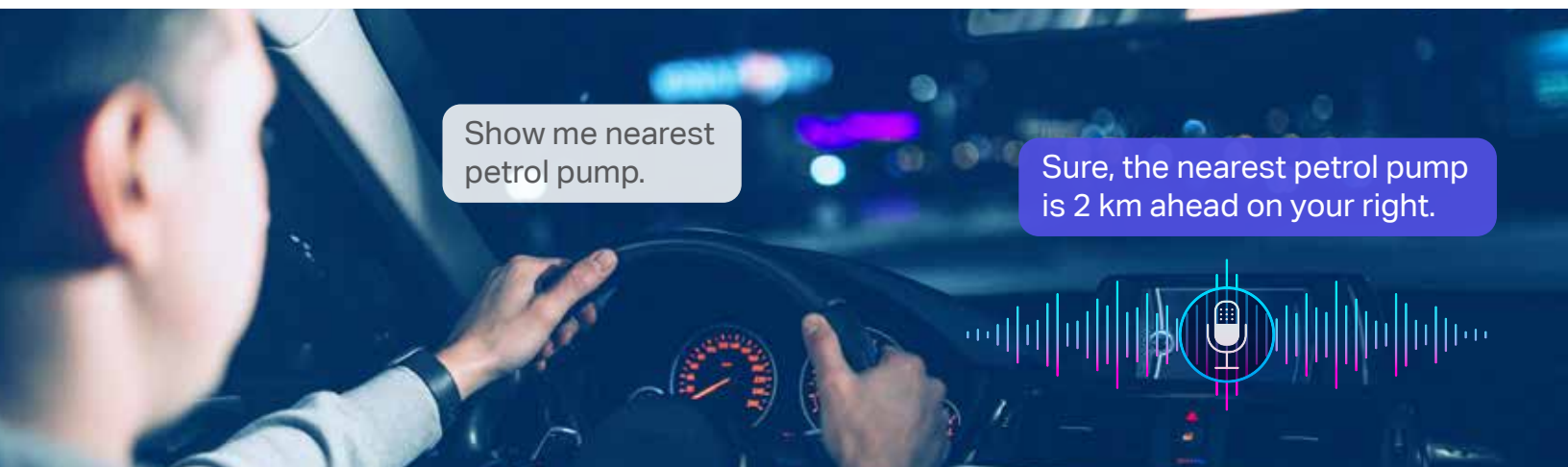
SOLUTION

Shaip with its expertise in the Conversational AI space enabled the client with:

- » **Data Collection:** Over **200k** key phrases or brand prompts collected in **12** global languages such as Norwegian, Danish, Swedish, Czech, Hebrew, UK English, French, Italian, Polish, Arabic, Turkish, UK English, India – English from **2800** speakers spread all across the globe.

Each language segment comprises a certain number of **native speakers** who provided **verbal recordings**. These recordings captured select key phrases such as **OK TOYOTA, HEY SALA, Hello Togg**, and many more such brand key phrases as required by the client. The collected data **varies in terms of the number of phrases used per speaker and the frequency of these recordings**. Notably, the English language in one Asian country yielded the most extensive set of data, given its larger speaker base and diversity in key phrases.

- » **Diverse Accents & Dialects:** Shaip effectively recruited specialists from around the world, proficient in the desired accents and dialects.
- » **Context & Intent Recognition:** Each speaker recorded key phrases in 20 variations, enhancing the ML models' understanding of context and intent.
- » **Background Noise Handling:** To ensure clear audio quality in real-world settings with background noises, we recorded key phrases in quiet environments with noise levels $\leq 40\text{dB}$, free from disturbances like TV, radio, music, speech, or street sounds.
- » **Reducing Bias:** To reduce bias, we involved participants from various regions, ensuring a balanced 50% male and 50% female representation across ages 18 to 60.
- » **Recording Guidelines:** Key phrases were recorded in a consistent, normal speech pace without variations. Every recording had a 2-second silence at the start and end to ensure no speech was unintentionally clipped.
- » **Recording Format:** The audio was recorded at 16kHz, 16-bit PCM in mono, utilizing a single channel, and saved in the WAV file format. The audio remains unprocessed, meaning there was no application of compression, reverb, or EQ.
- » **Quality:** Shaip and the client jointly set the Quality Standards. Each speech recording underwent strict quality checks. Only those that met the criteria were delivered. Recordings not meeting the standards were re-recorded at no additional cost.



THE OUTCOME

The high-quality brand key phrase audio data or voice prompts will enable the automotive companies and their customers with:



Branding and Identity: Voice prompts with specific, branded phrase helps the companies create a direct and memorable connection between the user and the brand that enhances brand recall and loyalty.



Ease of Use: Voice commands make it easier for drivers to interact with their vehicle's systems without taking their hands off the wheel or their eyes off the road thereby enhancing road safety and user experience.



Functionality: Voice commands make accessing and controlling car features more intuitive. Whether its navigation, media playback, climate control, or accessing real-time data, voice prompts simplify the interaction.



Integration with Other Systems: Many voice-activated systems integrate with smartphones, & smart home devices. E.g., a user could request their car to switch on the home lights as they near, or inquire about the weather at their destination.



Competitive Advantage: Offering advanced voice-activated systems can be a selling point and a differentiator in a competitive market. Buyers often look for the latest tech features when considering a new car purchase.



Future-Proofing: As technology advances and the (IoT) becomes more ingrained in daily life, a strong voice-activated system allows automotive companies to be better prepared for future tech innovations.

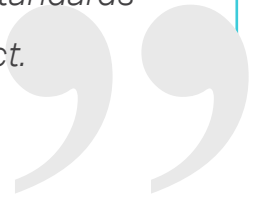


Revenue Opportunities: Additional monetization opportunities exist. For example, voice systems might suggest integrated e-commerce experiences, such as ordering food or locating nearby services, leading to potential affiliate revenue.

CUSTOMER TESTIMONIAL



When we began sourcing voice prompts for the automotive sector, the challenges were numerous. Capturing the diversity in speech, accents, and tones was vital to represent our client's global clientele. Shaip stood out not just as a vendor, but as a true partner. Their commitment to securing a diverse range of voices from different regions was commendable. They went beyond merely gathering voices; they grasped the nuances of our project needs, guaranteeing top-notch recordings. Their flawless adherence to audio collection standards showcased their professionalism and dedication to the project.



Headquartered in Louisville, Kentucky, Shaip is a fully managed data platform designed for companies looking to solve their most demanding AI challenges enabling smarter, faster, and better results. Shaip supports all aspects of AI training data from data collection, licensing, labeling, transcribing, and de-identifying by seamless scaling of our people, platform, & processes to develop AI/ML models. To learn more about how to make your data science team and leaders' life more manageable, visit us at www.shaip.com.

SoundHound

SoundHound AI (Nasdaq: SOUN), a global leader in conversational intelligence, offers voice AI solutions that let businesses offer incredible conversational experiences to their customers. Built on proprietary technology, SoundHound's voice AI delivers best-in-class speed and accuracy in numerous languages to product creators across automotive, TV, and IoT, and to customer service industries via groundbreaking AI-driven products like Smart Answering, Smart Ordering, and Dynamic Interaction™, a real-time, multimodal customer service interface. www.soundhound.com