

CASE STUDY

Training data to build multi-lingual Conversational AI

High-quality audio data sourced, annotated, and transcribed to train Conversational AI in 40 vernacular languages in less than 30 weeks



COMPANY

A leading conglomerate



USE CASE

Conversational AI



OUR OFFERING

- Audio Sourcing
- Audio Annotation
- Transcription

KEY STATS

Hours of Audio Data
Sourced, Annotated
& Transcribed

20,600+

No. of
Languages

40

Project
Duration

30 WEEKS

(2x-5x faster than the competition)

OVERVIEW



Chatbots and Digital Assistants are based on the underlying technology of Conversational AI. Unfortunately, these virtual assistants are only as smart as the technology and data that was used to train them. The data lets you bring your AI to life by automating and streamlining activities, improving enterprise productivity, and boosts customer engagement. Chatbots and Digital Assistants are subsets of Conversational AI.

Conversational AI uses a set of technologies ranging from natural language processing (NLP) to real-time machine learning (ML), that enable machines to simulate conversations just like humans. A conversational AI needs data for two main reasons: to know what people are saying, and what it needs to respond.

CHALLENGES

Lack of quality training data related to conversational AI has been a bottleneck in its progress & adoption. Sourcing audio, annotating, and transcribing datasets in large volumes with the required quality, sufficient enough to build AI capabilities, has always been a time-consuming, expensive task that requires skilled resources from various domains. The critical requirements of the client were:



Data Sourcing

- Acquire hours of conversational audio data in different languages on a range of topics & variety of media domains in 8kHz & 16kHz sampling rate.
- Ensure diversity in datasets - domains, speakers demographics, background etc. to train Conversational AI in an unbiased way.



Data Transcription

Complex audio annotation & transcription guidelines with min. accuracy of 95% for word accuracy & 90% for tag accuracy.



Data Delivery

Transcript files to be delivered in JSON format as per specifications within 30 weeks.

SOLUTION

With our deep understanding of conversational AI, we helped the client source, annotate, transcribe the right set of audio data required to train their AI-enabled speech model, with utmost precision. We offered audio data collection and transcription services based on their requirements while fully customizing desired intent, utterances, and demographic distribution. We helped the client with:

1. Audio Sourcing

A. 8 kHz Data

Acquired **9,900+** hours of unbiased/unscripted quality audio data (Call Center / General Conversation) on a range of 17 general topics i.e. Finance, Insurance, Retail, Telecom, Hospitality, Legal, Family, Friends, Culture, Sports, Work, Study/School, Health, Movies/TV Shows, Information & Technology, News & Current Affairs, Holiday / Leisure, Agriculture, Aviation, Movies, Oil and Energy, Real estate, etc.

B. 16 kHz Data

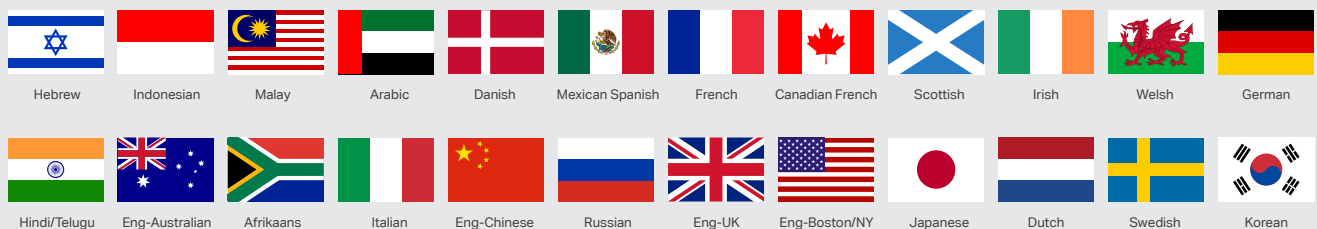
Acquired **10,800+** hours of high-quality audio data from a wide variety of media domains: Arts and Culture, Beauty and Lifestyles, Biography, Cars and Motors, Consumer and Retail, Crime and Justice, Entertainment and Music, Family and Relationship, Food and Drink, Sports, News, Broadcast, Talk / TV Shows, etc. from a diverse set of speakers with respect to their accents, gender, and demographics.

C. Total Data

Acquired over **20,600+** hours of high-quality audio data across 40 different languages in multiple dialects from over **3,000+** experienced and credentialed linguists across the world, so as to train the Conversational AI agent in an unbiased way.

Languages Supported

Afrikaans, Arabic, Bengali, Chinese Mandarin, Danish, Hebrew, Hindi, Indonesian, Kannada, Korean, Malay, Marathi, Polish, Swahili, Swedish, Telugu, Thai, Vietnamese, New Zealand English, South African English, Hindi - English (Hinglish), Singaporean English (Singlish), Indian English, Australian English, UK English, US Spanish, US English, African-American Vernacular English, Hispanic English, Chinese English, Boston English, Newyork English, Deep South English, Irish English, Scottish English etc.



& many more...

2. Transcription

For the Audio data collected, the annotation and transcription met the minimum accuracy scores set by the client i.e., passing validation **95%** for word accuracy and **90%** for tag accuracy.

The team also successfully adhered to strict & complex guidelines for transcriptions set under two broad categories apart from the general guidelines i.e., 1) Speech event transcription & 2) Non-speech (acoustic event) transcription guidelines.

A. Speech event transcription guidelines

- 1.1. Use orthographic spelling (Contractions, Abbreviations, Stumbled speech and corrections, Filler words, Interjections, overlapping speech, Letters spoken as letters)
- 1.2. Punctuation (Commas, Exclamation marks, Apostrophes, Hyphens, Tildes, Special symbols)
- 1.3. Capitalization, Numbers, Acronyms and initialisms, Unintelligible words and phrases, Multiple Languages (code-switching)

B. Non-speech (acoustic event) transcription

- 1.1. Non-speech sound inventory
- 1.2. No Speech
- 1.3. Music only

C. Data Delivery

All audio and transcript files were delivered in. JSON format in accordance with the specified metadata requirements within the stipulated time.

THE OUTCOME

The high-quality audio data empowered the client to train its Conversational AI on a wide variety of topics, ranging from Telecom, Hospitality to Legal in 40 different languages and dialects to mimic human conversation. The benefits that the client derived from the platform were:

- It can seamlessly interact with humans in multiple languages.
- It can learn from existing interactions and can suggest & converse, based on these learnings
- It can understand the intent behind the human speech and remove ambiguity in understanding human language
- There are 1,000's of use cases that can be achieved; few of them are: Book a ride, Play Music, seek an answer to a query, order office supplies, order food, get updates to latest news, check the weather, live score, stock prices, Check Flight Prices, Hand Free Text Messages, etc.

CUSTOMER TESTIMONIAL



"We were pleasantly surprised with the quick turnaround time and Shaip's domain expertise to source, annotate, and transcribe high-quality audio files in 26 different languages. The highly customized data has enabled us to newer geographies quickly."



Headquartered in Louisville, Kentucky, Shaip offers a human-in-the-loop platform and services to support all aspects of managing training data for the development of AI/ML models. From data collection, licensing, curation, labeling to the seamless scalability of our people, platform, and processes, Shaip contributes to a diverse set of verticals. To learn more, visit us at www.shaip.com.