



GCP の高精度音声認識・翻訳技術が支える 手のひらサイズの翻訳機



ソースネクスト株式会社

<https://sourcnext.co.jp/>
〒105-7133 東京都港区東新橋 1-5-2
汐留シティセンター 33 階

1996年8月、ソフトウェアの企画・開発・販売を事業として設立。その後、スマホアプリに事業を拡大し、さらに現在は、ポケットークをはじめとするIoTデバイスの企画・開発・販売を推進。コンシューマ向け製品を中心に、数百種類のアイテムを、家電量販店、自社ECサイト、携帯キャリアの3つの販路により展開。登録数約1,700万人のユーザーデータを活用した顧客サービスの向上にも取り組んでいます。

執行役員 技術戦略室
シニア製品設計スペシャリスト
川竹 一 氏

「世界一エキサイティングな企業になる」というビジョンに基づき、製品を通じて喜びと感動を世界中の人々に広げることを目指すソースネクスト株式会社（以下、ソースネクスト）。2019年12月に累計出荷台数が60万台を突破した、ボタンを押しながら話しかけるだけで通訳がいるかのように対話ができるAI翻訳機「POCKETALK®（ポケットーク）」（以下、ポケットーク）の開発について、ソースネクスト株式会社 執行役員 技術戦略室 シニア製品設計スペシャリスト 川竹一氏に話を伺いました。

(利用している Google Cloud Platform サービス)
[Cloud Speech-to-Text](#)、[Cloud Text-to-Speech](#)、[Cloud Translation API](#)

音声認識や翻訳エンジンはGCPを採用、自社ではUI開発に注力し 短期間での商品化を実現

「ソースネクストは、海外のソフトウェアをローカライズして、国内で販売するビジネスからスタートした会社です。海外でのビジネスに英語が必須でしたが、言語の壁は高く、その壁がなくなれば、より一層ビジネスがやりやすくなると感じていました。言語の壁をなくす方法は2つ。1つは人が言語を学習すること、もう1つは機械翻訳を利用することです。そこで我々は翻訳機を作ろうということになりました。2001年ごろのことです」と話すのは、執行役員 技術戦略室 シニア製品設計スペシャリストの川竹一さんです。

しかし当時のネットワーク環境は、ブロードバンドがスタートしたばかりの状況で、翻訳ソフトや音声認識技術もPCでしか利用できませんでした。リアルなコミュニケーションの場で利用できる翻訳機の本格開発は、その後のスマートフォンの登場と高速ネットワーク環境の進化を待つことになりました。

2017年10月、グローバルで第4世代移动通信システム(4G)が利用できるようになり、ディープラーニングの実用化により、音声認識や翻訳の精度も向上してきた背景の後押しも受け、初代ポケットークが誕生します。初代のポケットークは非常に評判がよく、初回生産台数分は

予約開始から約11日で完売。すぐに再販を決めますが、その間にも数多くのフィードバックが寄せられました。これらの要望に対応するため、開発から製造までを自社で行うことに切り替え、2018年9月、2代目となるポケットークWを誕生させます。開発のコンセプトは、「母国語の違う初対面の話者同士が簡単にコミュニケーションできるツール」です。

このポケットークWの開発プラットフォームとして、Google Cloud Platform(GCP)の機械学習による発話からテキストへの変換サービスであるCloud Speech-to-Text、および高速かつリアルタイムの翻訳サービスであるCloud Translation APIが採用されています。

Cloud Speech-to-Textが採用された理由は、120の言語と方言を高い精度で認識できること。一方、Cloud Translation APIが採用されたのは、100以上の言語の翻訳を高い精度で実現できることが理由でした。川竹さんは、「加えて、GCPを使い慣れていて、必要なドキュメントがそろっていたことなども採用した理由の1つでした」と話しています。

まず重視したのが、ポケットークの持ち主はもちろん、初めて使う相手

にも分かりやすい簡単な操作性を実現することです。「IT リテラシーがあまり高い利用者でも、箱から出してすぐ使えることを目指しました」(川竹さん)。ポケット W にはグローバル通信機能が内蔵され、別途 SIM を購入することなく、世界中のどこでも利用できるような工夫もされています。利用者は 50 代～ 70



1 か月間で 101 の国と地域で利用されている - 専用グローバル SIM 利用実績 (日本からの利用を除く) を元にソースネクストによる集計 (2018 年 8 月)

代がボリュームゾーンで、海外旅行で利用されるケースが多くなっています。また、インバウンドの増加を受け、日本国内でのさまざまな店舗における外国人旅行者への対応や、企業内の外国人労働者とのコミュニケーションなどにも利用が拡大しているといえます。

「ポケット W の翻訳のしくみそのものはとてもシンプルですが、初めて出会った人同士がストレスなく会話できるスピードや、翻訳ボタンのみで直感的に使えるインタフェースなど、クリアしなければならない条件がありました。特に、音声認識や翻訳エンジンを一から開発すると、膨大な時間と工数が必要です。GCP を利用することで、音声認識や翻訳は GCP に任せ、スピードやインタフェースなど、実用面の工夫に時間と工数を集中することができました。」(川竹さん) 累計出荷台数はすでに 60 万台を超え、個人のみならず、企業での利用も増えつつあります。「同様の翻訳機は他社からも提供されていますが、ポケット W の強みは、音声認識の精度が高いことです」と話す川竹さん。本体にノイズキャンセラーを搭載することで、周囲の雑音に影響されることのないコミュニケーションを実現しています。今後は、国内だけでなく、海外の販路も拡大していく予定。2019 年中には、世界 20 か国以上に展開する計画です。

高精度の音声認識と翻訳、広がる利用シーン

ポケット W の開発では、2017 年末より音声認識や翻訳エンジンの実験を開始し、2018 年初めより翻訳機のプロトタイプを作成を開始。Android デバイスでテストを繰り返し、約 9 か月という短期間で開発しています。川竹さんは、「GCP による開発は、非常に楽でした。開発に携わったエンジニアが GCP を使い慣れていたこともありすが、すごいスピードでサンプルプログラムを作ってくれました」と話します。ポケット W の開発に GCP を採用した最大の理由は、音声認識や翻訳の精度でしたが、高速なレスポンスも採用理由でした。また、言語のカバレッジも広く、少ないコストで試すことができることもメリットの 1 つでした。プロトタイプ終了後に、いくつかの音声認識や翻訳エンジンを比較検討した結果、GCP の採用を決めています。川竹さんは、「特に音声認識に関しては、GCP はバツグンに高い評価でした」と話しています。

システムの基本構成は、音声認識、翻訳、音声合成の 3 ステップ。ポケット W で認識した音声を Cloud Speech-to-Text により音声認識処理をしてテキスト化し、このテキストを Cloud Translation API で、あらかじめ設定された言語に翻訳、Cloud Text-to-Speech により、言語ごとの音声に変換してポケット W に返す仕組みです。複数社のエンジンをミックスして利用しており、言語対や利用状況により最適なエンジンに切り替えることで、幅広い地域や言語で高い精度を確保しています。

音声を認識し、翻訳して、音声合成をするのは、リレーのような仕組

みです。そのため、トップランナーでもある音声認識が、音声を正しく認識できないと、その後の翻訳の精度も低下してしまいます。川竹さんは、「GCP は、音声認識の精度が非常に高いので、結果として高い翻訳精度を実現できました。常に 1 番いいテクノロジーを使うのが、システム開発におけるポリシーです」と話します。

ポケット W は、すでに交通機関や宿泊施設、小売店、飲食店、工場、自治体、教育機関など、さまざまな業種に利用の場を拡大しつつあります。川竹さんは、「ある EC サイトのユーザーレビューでは、“ポケット W に命を救われました” という書き込みもありました。友人が日本で転んでけがをしたときに、病院でポケット W を使うことで適切な処置をしてもらえたことができたということでした」と話します。

「正直、命にかかわることもある医療機関では使われないだろうと思っていたのですが、お医者さんは、日本語が話せない患者さんの来院時、何に困っているのかが分からず苦勞しているということでした。生活のあらゆるシーンで、ポケット W が役立つ場面が増えていると感じています。」(川竹さん)

今後の取り組みを川竹さんは、「ユーザーのフィードバックに基づいた改善をさらに進めていきたいと思っています。その一環として、レスポンスの向上を目的に、サービス全体を GCP に移行することも検討しています。今後もポケット W のようなヒット商品を生み出すべく、IoT デバイスの開発に注力していきます。そのためのサポートを、Google Cloud には期待しています」と話しています。

Google Cloud を活用することで、ビジネスの将来に注力できるようになります。インフラストラクチャの管理やサーバーのプロビジョニング、ネットワークの構成などに起因する負担を軽減することができます。つまり、イノベーターもプログラマーも、自分の本来の仕事に集中することができます。

お問い合わせはこちら
<https://goo.gl/CCZL78>

