



## FUJITSU Hybrid IT Service FJcloud-0 Voice Operation



機能概要

---

# 本書をお読みになる前に

## ご注意

- 本書の著作権は、富士通株式会社にあります。
- 本書の一部または全部を無断で使用、複製することはできません。
- 本書に記載されている事柄は、将来予告なしに変更することがあります。
- 本書に記載されているシステム名、製品名等には、必ずしも商標表示 (™, ®) を付記していません。
- その他、記載されている会社名、製品名は、各社の商標または登録商標です。

---

# 目次

1.	はじめに .....	3
1.1.	本書の目的 .....	3
1.2.	本書の読者 .....	3
1.3.	本書の構成 .....	3
1.4.	用語解説 .....	3
2.	提供機能 .....	3
2.1.	提供する SDK .....	3
2.2.	機能一覧 .....	3
3.	各機能概要 .....	5
3.1.	基本機能 .....	5
3.1.1.	音声認識サーバ情報取得 .....	5
3.1.2.	音声認識開始 .....	5
3.1.3.	音量補正 .....	6
3.1.4.	優先単語辞書 .....	6
3.1.5.	音声認識結果通知 .....	6
3.1.6.	音声認識経過結果通知 .....	7
3.1.7.	音声認識停止（キャンセル） .....	7
3.1.8.	プロキシ設定 .....	7
3.2.	オプション API .....	8
3.2.1.	音量レベル登録 .....	8
3.2.2.	フレームサイズ設定 .....	8
3.2.3.	有声検知レベル閾値設定 .....	8
3.2.4.	無声判定時間設定 .....	8
3.2.5.	自動発話開始タイムアウト時間設定 .....	8
3.2.6.	連続認識設定 .....	8
3.2.7.	最大発話時間設定 .....	9

---

# 1. はじめに

## 1.1. 本書の目的

本書は、FUJITSU Hybrid IT Service FJcloud-O Voice Operation（以下、Voice Operation と記載）が提供する機能の概要について記載します。

機能に関する詳細は、Voice Operation SDK（以下、SDK と記載）に含まれる「API リファレンス」および「アプリケーション開発ガイド」をご参照ください。

## 1.2. 本書の読者

本書は、Voice Operation の利用を検討されている方、概要を知りたい方を対象としています。本書を読むにあたっては、以下の知識が必要です。

- 端末に関する基本的な知識

## 1.3. 本書の構成

本書は、次の構成になっています。

1. はじめに  
本書の目的・概要を説明します。
2. 提供機能  
提供している SDK と機能の一覧を説明します。
3. 各機能概要  
機能ごとの概要を説明します。

## 1.4. 用語解説

本書で使用している用語について、表 1-1 で説明します。

表 1-1 用語解説

名称	本書での表記	説明
Distributed Speech Recognition	DSR	クラウド型音声認識。音声認識サーバに音声データを送信し、音声認識を行う形態のこと。
Voice Operation API	API	Voice Operation の各種機能にアクセスするため用意された、クライアント API のこと。
Voice Operation SDK	SDK	Voice Operation の各種機能を利用したアプリケーションを開発するために必要なソフトウェア開発キットのこと。
Voice Operation アプリケーションサーバ	アプリケーションサーバ	Voice Operation の各種機能を提供するアプリケーションサーバのこと。
Voice Operation 音声認識サーバ	音声認識サーバ	音声認識エンジンを搭載したサーバのこと。音声認識エンジンは、入力された音声データをテキストデータに変換する機能を持つ。
音量補正值	音量補正值	マイクから入力された発話の音量を補正する値。発話された音量を 1% から 1000% (0.01 倍から 10 倍) の範囲で補正することが可能。
音量レベル	音量レベル	音量補正值によって補正された音量をデシベル換算し、0 から 100 の範囲で表したものの。
音声認識結果	音声認識結果	発話終了後に音声認識サーバから返ってくる結果 (複数候補の場合有)
音声認識経過結果	音声認識経過結果	発話中に音声認識サーバから返ってくる結果 (複数候補の場合有)
優先単語辞書	優先単語辞書	ユーザが優先的に認識させたい語彙を登録する端末内の辞書。音声認識時に辞書を使用することで登録した語彙の出現率を上げることが可能。

## 2. 提供機能

本章では、SDK の種類および SDK が提供する機能について説明します。

### 2.1. 提供する SDK

Voice Operation では、以下に記載の OS に対応した SDK を提供しています。

- Android
- iOS
- Windows デスクトップ

それぞれの SDK には、表 2-1 のファイルが含まれています。

SDK は、Voice Operation ポータルより入手することができます。

表 2-1 SDK の内容物

内容物	説明
アプリケーション開発ガイド	アプリケーション設計・開発を行う際に必要な SDK の使用方法を記載しています。
API リファレンス	提供する API の詳細仕様を記載しています。
ライブラリ	API を使用するためのライブラリ群
サンプルアプリケーション	API を使用した各 OS のサンプルアプリケーション

### 2.2. 機能一覧

SDK で提供する機能の一覧と、各 OS の対応を表 2-2 で説明します。

オプション機能とは、音声認識の精度をより向上させるための機能群です。

表 2-2 機能一覧と OS の対応表

提供機能		Android	iOS	Windows デスクトップ
基本機能	音声認識サーバ情報取得	○	○	○
	音声認識	○	○	○
	音声認識結果通知	○	○	○
	音量補正	○	○	○
	優先単語辞書	○	○	○
	音声認識経過通知	○	○	○
	音声認識停止（キャンセル）	○	○	○
	プロキシ設定	○※1	○※1	○※1
オプション機能	音量レベル登録	○	○	○
	フレームサイズ設定	○	○	○
	有声判定閾値設定	○	×	○
	終話検知時間設定	○	○	○
	自動発話開始タイムアウト 時間設定	○	×	×
	連続認識設定	×	×	○
	最大発話時間設定	○	○	○

※1 OS によって設定できる内容が異なります。（3.1.8 章）

## 3. 各機能概要

本章では、2章で記載した各機能の概要を説明します。

### 3.1. 基本機能

#### 3.1.1. 音声認識サーバ情報取得

音声認識処理時に接続を行う音声認識サーバへの接続情報をアプリケーションサーバより取得します。本機能はアプリケーションの起動後、音声認識処理より前に必ず実行する必要があります。本機能を実行するためには、以下のお客様固有の情報が必要です。それぞれの情報は Voice Operation ポータルで確認することができます。

- アプリケーションサーバの URL
- 法人 ID
- テナント ID
- API キー

#### 3.1.2. 音声認識開始

音声認識を開始します。本機能を実行するためには、あらかじめ音声認識サーバ情報取得機能で音声認識サーバへの接続情報を取得しておく必要があります。

音声認識では開始時に、音量補正（3.1.3章）、優先単語辞書の使用有無（3.1.4章）また音声認識経過結果の取得有無（3.1.6章）を指定することができます。

音声認識を開始すると端末は音声認識中となり、ユーザの発話を受け付けます。受け付けた音声データは、音声認識サーバに送信します。また、音声認識中には処理状態に応じたコールバックをアプリケーションに通知します。コールバックを表 3-1 で説明します。

表 3-1 音声認識中のコールバック

パラメータ名	説明
音声認識サーバ接続完了通知	音声認識サーバとの接続が完了したことを通知します。
音声認識開始処理完了通知	音声認識が正常に開始されたことを通知します。
発話検知通知	ユーザの発話を検知したことを通知します。

パラメータ名	説明
音量レベル通知	音声認識中の音量レベルを一定間隔で通知します。
音声認識経過結果通知	音声認識中の認識結果を一定間隔で通知します。(3.1.6章)
終話検知通知	ユーザの発話終了を検知したことを通知します。
音声認識結果通知	音声認識結果を通知します。(3.1.5章)
異常通知	音声認識が異常終了したことを通知します。

### 3.1.3. 音量補正

音声認識時にマイクから入力された音の音量を補正する機能です。指定された音量補正值にしたがって、発話された音量を 1%から 1000% (0.01 倍から 10 倍) の範囲で補正することができます。音量補正值の指定には、SDK が自動で音量補正值を算出する自動補正と、ユーザが音量補正值を直接指定する補正值指定の 2 種類があり、音声認識時のパラメータで指定することができます。

### 3.1.4. 優先単語辞書

ユーザが優先的に認識させたい単語を登録した端末内の辞書です。辞書は音声認識開始時にサーバに送信されます。

音声認識時に辞書を使用する事で、Voice Operation の言語モデルに存在しない単語を追加したり、変換候補の出現率を上げたりすることができます。辞書の使用有無は音声認識時のパラメータで指定することができます。辞書の作成・編集は API を使用して行うことができます。

### 3.1.5. 音声認識結果通知

音声認識の結果通知では、以下の情報を取得することができます。

- 音声認識結果テキスト
- 音声認識詳細結果

音声認識結果テキストとは、音声認識で得られる単語毎の変換候補から、単語毎の第一候補を結合した文章です。

音声認識詳細結果とは、単語毎の変換候補を全て格納したリストです。単語毎の変換候補毎に以下の内容が含まれています。

- 表記
- ヨミ
- 品詞（日本語の言語モデルのみ）
- 発話開始時間（日本語の言語モデルのみ）
- 発話終了時間（日本語の言語モデルのみ）

### 3.1.6. 音声認識経過結果通知

音声認識中の経過結果を通知する機能です。音声認識中に経過結果が一定間隔で通知されます。通知される間隔は設定されているフレームサイズ（ms）に依存します。（3.2.2 章）認識経過通知で取得できる情報は、音声認識結果通知で取得できる情報と同じです。（3.1.5 章）認識経過結果の取得有無は音声認識時のパラメータで指定することができます。

### 3.1.7. 音声認識停止（キャンセル）

音声認識を停止する機能です。停止するとそれまでに発話した音声認識結果は破棄されます。

### 3.1.8. プロキシ設定

プロキシサーバを経由して Voice Operation サーバと通信を行う機能です。対応しているプロキシ認証方式は Basic 認証のみです。API でプロキシサーバの情報を指定することができます。指定できる内容は OS によって異なります。OS 毎に設定可能な内容を表 3-2 で説明します。プロキシを指定しない場合には、端末のネットワーク設定にしたがって動作します。

表 3-2 OS 毎に設定可能なプロキシ情報

パラメータ名	Android	iOS	Windows デスクトップ
プロキシホスト	○	×	○
プロキシポート番号	○	×	○
認証ユーザ名	○	○※	○
認証ユーザ名	○	○※	○

※ 音声認識サーバとの通信時に使用されます。

## 3.2. オプション API

### 3.2.1. 音量レベル登録

音量レベルをアプリケーションサーバに登録します。音量レベルに登録する事で、アプリケーションの再起動後でも自動補正による音量補正を再起動前と同じレベルに調整することができま  
す。登録を行わない場合は、アプリケーション再起動後の自動補正による音量補正は規定値に戻  
ります。

### 3.2.2. フレームサイズ設定

音声認識で使用する音声入力バッファのサイズをミリ秒単位で設定します。音声認識中には、  
フレームサイズの設定間隔ごとに音量レベル通知と音声認識経過結果通知が通知されます。小さ  
な値に設定するほど通知間隔は短くなりますが、アプリの処理負荷が上がり音声認識の性能にも  
影響が出る場合があります。

### 3.2.3. 有声検知レベル閾値設定

音声認識時にユーザの発話を検知する際の有声検知レベルの閾値を設定します。小さな値に設  
定するほど小さな声でも発話検知しやすくなりますが、周囲の会話などの環境音でも発話検知し  
やすくなります。

### 3.2.4. 無声判定時間設定

音声認識時の無声判定時間をミリ秒単位で設定します。発話検知後は、設定された無声判定時  
間を超える無声区間があると終話と判断し音声認識を終了します。短い時間に設定するほど発話  
の終了から認識結果を受けるまでの時間も短くなりますが、ちょっとした言いよどみなどでも終  
話と判断されやすくなります。

### 3.2.5. 自動発話開始タイムアウト時間設定

発話が始まるまでのタイムアウト時間をミリ秒単位で設定します。音声認識が開始されて  
から発話検知を行うまでに、設定された自動発話開始タイムアウト時間の設定値を超えた場合、  
音声認識を終了します。

### 3.2.6. 連続認識設定

連続認識を有効にするかどうかを設定します。音声認識前に有効に設定しておくことで、音声

認識の完了後に SDK が音声認識を繰り返し実行します。繰り返し実行される音声認識では、通常の音声認識と同じイベント、同じ結果を受け取ることができます。連続認識を停止させる際は音声認識停止（キャンセル）を実行します。（3.1.7 章）

### 3.2.7. 最大発話時間設定

一度の音声認識で発話可能な最大発話時間を秒単位で設定します。発話を検知してから設定された最大発話時間を超えた場合、音声認識は自動終了します。最大発話時間を超えて自動終了した場合は、それまでの発話に対する音声認識結果は受け取ることができます。最大発話時間は、既定では 10 秒に設定されており、最大で 30 秒まで設定することができます。

本設定値は SDK では変更することができません。本設定値を変更可能なお客様は音声認識サービス（専有環境）をご利用のお客様に限られます。

本設定値の変更をご希望のお客様はヘルプデスクサービスより個別にお申し込みください。

---

FUJITSU Hybrid IT Service FJcloud-O Voice Operation  
機能概要

©FUJITSU LIMITED 2016-2020

発行日 2020年09月  
発行者 富士通株式会社

〒105-7123 東京都港区東新橋 1-5-2 汐留シティセンター

---