

雑音に強く高認識率の音声認識システム

Voice Magic® シリーズ



1

概要

レイトロンが独自に開発した、雑音に強く高認識率の音声認識エンジン「VoiceMagic」※1は、生活雑音環境下における音声認識を可能にしています。これまでの音声認識と大きく違う点は、雑音の多い実生活環境下においてマイクから2~3メートル離れた所から音声認識ができることです。また、音声認識をする際に、スイッチを押す必要が無い Always Listening 機能により、完全ハンズフリーを実現しています。通信回線が不要で、ネットワークに接続する必要が無いため、常時リアルタイム応答が可能です。「VoiceMagic」は様々な製品への展開が可能で、情報家電、白物家電、自動車、医療・バイオ、産業機器、介護ロボットなど幅広く活用することができます。

2

特徴

- 雑音環境下における高い認識率(SNR20dBで認識率98%以上)により、マイクから離れた所からの認識が可能
- 高性能な自動音声区間検出※2による、Always Listening 機能搭載
音声認識時にスイッチ操作や特定ワード、特定動作(手を叩く等)の認識トリガーが不要であるため、完全なハンズフリーシステムを実現可能
- 自動認識棄却フィルタリング※2により、登録フレーズ以外の入力音声に対する誤認識を防止し、認識精度を向上
- 話者の音声登録が不要(不特定話者対応)
- シーン機能により、登録フレーズをシーン別(最大16シーン)に分類することで類似単語への対応が可能
- 音声認識専用LSIを搭載したハードウェアシステムにより、搭載機器には高性能なCPUやDSPが不要
- 多彩な入力音声フォーマットに対応(I²S出力に加えPDM出力のMEMSマイクに対応可能)

3

製品への展開例

「VoiceMagic」は機器操作に適した音声認識システムであり、様々な機器に対して音声によるハンズフリー操作を提供します。

「VoiceMagic」を搭載することで音声による機器操作が可能

組込み型、後付け型、無線型など用途に応じて最適なタイプを選択

Voice Magic® Standard Voice Magic® USB Voice Magic® Bluetooth





Voice Magic[®]

機器操作に適した音声認識システム

4

製品ラインナップ

「VoiceMagic」は様々な用途、環境で活用できるように以下の製品ラインナップを揃えております。

製品名	外観 / 寸法	導入時の要件
Voice Magic[®] USB 〈後付け※1・USB型〉	 〈サイズ〉 W:55mm D:16mm H:5mm ※突起部含む	【通信要件】 ・USBシリアルによる通信機能を有していること ・アプリケーションソフトに「VoiceMagic」との通信機能を追加※2 (「VoiceMagic」は認識結果に応じた識別番号を出力)
Voice Magic[®] Bluetooth 〈後付け※1・無線型〉	 〈サイズ〉 W:55mm D:16mm H:10mm ※突起部含む	【通信要件】 ・Bluetooth SPPIによる通信機能を有していること ・アプリケーションソフトに「VoiceMagic」との通信機能を追加※2 (「VoiceMagic」は認識結果に応じた識別番号を出力)
Voice Magic[®] Standard 〈組み込み型〉	 〈サイズ〉 W:35mm D:24mm H:9mm ※突起部含む	【通信要件】 ・使用しているマイコンがI ² C又はUART通信機能を有していること ・マイコンプログラムに「VoiceMagic」との通信機能を追加※2 (「VoiceMagic」は認識結果に応じた識別番号を出力)
Voice Magic[®] EVA 〈特殊用途〉	 〈サイズ〉 W:57mm D:44mm H:18mm ※突起部含む	アナログマイク入力やRS232C出力が必要な場合を対象とした製品であり、音声認識機能をFPGAで実現しているため、機能変更や拡張などの追加対応が可能。 ※追加対応には別途費用が必要となります。

※1 必ずしも組込む必要がないため、後付け可能。「VoiceMagic」との通信要件を満たしておくことでオプション機能として実現可能。

※2 「VoiceMagic」との通信に関するサンプルコードは提供可能。

5

音声認識HMM学習データのパッケージシリーズ

高認識率の音声認識を実現するために重要なHMM学習データに、各種用途に応じた基本フレーズを登録した、操作パッケージシリーズが登場しました。シリーズは、家電・車載機器・生産機器・パソコン・ロボットの計5種類、用途に応じてパッケージを組合せて使用することも可能です。



6

適応環境

「VoiceMagic」は生活雑音環境下で使用できるとともに、ウェアラブル端末に搭載することで工場などの強雑音環境下においても音声認識によるハンズフリー操作を実現可能。

生活雑音環境下 (リビングなど)

40dB~70dBの生活雑音環境下でマイクから2~3メートル離れて音声認識が可能。



強雑音環境下(工場など)

ウェアラブル端末に搭載することで70dB~90dBの強雑音環境下で音声認識が可能。





あらゆる生活シーンを私たち
“レイトロンの組込み技術”
が支えます。

音声認識システムは、オーディオ・家電・携帯電話・クルマ・工場・ロボットなど、さまざまなシーンやツールに、その技術は利用されています。レイトロンの技術が人と人のあいだをつなぎます。



解決します! 従来の課題を1問1答 FAQ

“家事作業中”や“車の運転中”も
音声認識が使えますか?

(課題:生活雑音環境下における認識性能の低さ)

生活雑音環境下においても
高い認識性能を実現

「VoiceMagic」はSNR5dBで90%以上の認識率が達成できるよう設計されており、生活雑音環境下にてマイクから2~3メートル離れて認識可能。マイク近傍であれば70~90dBの強雑音環境下にて認識可能。

発話したフレーズと異なる結果が
出力されませんか?

(課題:音声認識率が低く、誤認識が多い)

高い認識性能を実現するとともに
誤認識を低減

「VoiceMagic」はSNR20dBで98%以上の高認識率を実現するとともに棄却機能を搭載。これにより、Always Listening 機能搭載時の課題となる誤認識を低減し、機器操作時の誤動作を防止する。

ハンズフリーシステムを実現できますか?

(課題:ボタンを押す又は合言葉を言う等の
認識開始合図が必要)

いつでも音声認識が可能

「VoiceMagic」は自動区間検出による Always Listening 機能を搭載しており、ユーザーによるボタン操作や合言葉なしに音声認識が可能。これにより、完全なハンズフリーシステムが実現可能。

ヘッドセットなどの接話マイクを
使う必要はないですか?

(課題:接話マイクを使用しないと所望の
認識性能が得られない)

マイクから離れて音声認識可能

「VoiceMagic」はSNR5dBで認識できるよう設計されており、生活雑音環境下にてマイクから2~3メートル離れた所からの認識が可能。マイク近傍であれば70~90dBの強雑音環境下にて認識が可能。これにより、接話マイクやマイクロフォンアレイが不要となり、利便性向上に加え、コスト削減にも繋がる。

通信回線を使う必要はないですか?

(課題:大規模な音声認識用サーバーがないと
機能実現できない)

通信回線や音声認識用サーバーが不要

「VoiceMagic」は音声認識機能を専用ハードウェアで実現している。これにより、通信回線や音声認識用サーバーが不要となり、容易に機器に組み込むことができる。さらに通信回線費やサーバー管理費が不要となり、コスト削減にも繋がる。

認識フレーズの発話後すぐに
認識結果が得られますか?

(課題:応答速度が通信回線の状況に左右される)

常にリアルタイム応答が可能

「VoiceMagic」は音声認識機能を専用ハードウェアで実現している。これにより、使用環境に左右されず常にリアルタイム応答が可能。認識結果は発話終了後200ms程度で出力。



VoiceMagicの応用製品

● オーディオキャンセラー搭載音声認識システム

テレビやカーオーディオのスピーカーからの出力音声と生活雑音を混合した環境で認識可能。「VoiceMagic」前段にAV機器からの出力音声を利用したオーディオキャンセラーを搭載(適応型フィルター)。

● 音声認識コントロールボックス

「VoiceMagic」を搭載する事で生活雑音環境下における音声による各種家電操作を実現し、より快適な住環境を構築するバリアフリーなユーザーインターフェイスボックス。学習リモコンで各種家電機器等のリモコン信号を記憶させ赤外線を出力。さらに離れた場所では、無線I/F(ZigBee)で送信可能。

9

仕様

〈共通仕様〉

音声認識方式	雑音ロバスト孤立単語音声認識
話者	不特定話者対応 (事前登録不要)
言語	多言語対応可能 (日本語・英語・中国語・イタリア語など)
学習データ	独自フォーマット (HMM:Hidden Markov Model)
総登録フレーズ数	6000フレーズ (6000個のHMM※ ¹)
1シーンの最大登録フレーズ数	1000フレーズ (男性・女性・子供を認識する場合は300程度)
シーン数	16シーン※ ²
搭載機能	音声認識機能 / 音声区間自動検出機能 / 登録外単語棄却機能

〈LSI仕様〉

ホストインターフェイス (UART)	9600, 19200, 38400, 57600, 115200, 230400, 460800bpsから選択
ホストインターフェイス (I ² C)	最大400kbps, スレーブアドレスは4種から選択
入力音声フォーマット	左詰め, I ² S, PDM
外部メモリー	SPI Flashメモリー (最大128Mbit)
入力クロック	5.6448MHz, 11.2896MHz, 22.5792MHzから選択
電源電圧	コア 1.8V, I/O 3.3V

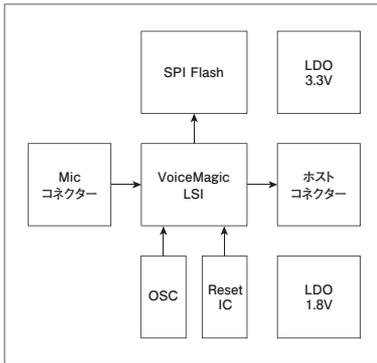
※¹ Flashメモリーに格納可能な学習データ数は最大6000個であり、男性・女性・子供を認識する場合は最大2000フレーズ登録可能。

※² 言語や用途に応じて認識フレーズセット(シーン)を変更可能。

10

ブロック図

Voice Magic[®] Standard



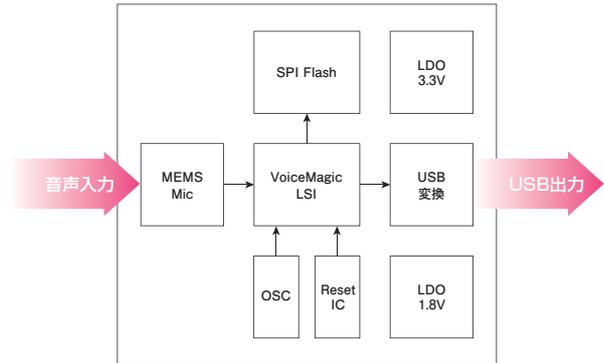
◆Mic コネクタ

- 1.電源 (3.3V)
- 2.M_BCLK
- 3.M_LRCLK
- 4.M_DAT
- 5.GND

◆ホストコネクタ

- 1.電源 (5.0V)
- 2.RX
- 3.TX
- 4.SCL
- 5.SDA
- 6.GND

Voice Magic[®] USB



11

認識動作及び性能

●認識処理概要



●音声認識エンジンの認識率シミュレーション結果[2]

ノイズ種類	測定条件	測定結果
雑音データ:NOISEX-92 白色雑音、ピンク雑音、HF、無線チャネル雑音、工場雑音、 人の声のざわめき、車内走行音、コクピット飛行音など	入力音声: 電子協音声データベース SNR:20dB	平均認識率:98.27%

[1] 【産学共同研究】音声認識の技術をさらに進化させるため、大学と共同研究を行っています。〈音声認識機能〉北海道大学大学院 宮永研究室

[2] S.Yoshizawa, Y.Miyanaga and N.Wada: "A Low Power VLSI Design of a HMM Based Speech Recognition System", Proceedings of IEEE Mid-West Symposium on Circuits and Systems, 2, 1:489-492(2002)より

※内容は予告なく変更することがありますので、あらかじめご了承ください。

RayTron[®] 株式会社レイترون

<http://www.raytron.co.jp>

〈お問合せ〉 営業部

〒541-0053 大阪市中央区本町1-4-8 エスリードビル本町11階

Tel.06-6125-0500 Fax.06-6125-0588 E-mail:support@www.raytron.co.jp