

SR-MOD100C 不特定話者認識フレーズ(SIフレーズ) カスタム登録の手引き

(C)2017 マイクロテクニカ

■概要

SR-MOD100Cは音声認識を行ってその結果をUART信号（非同期式シリアル信号）で出力するモジュールですが、本書ではこのモジュールに不特定話者認識フレーズ(以下SIフレーズと記載)を追加するための申請方法をご案内しています。

SIフレーズはあらかじめ決められたものがSR-MOD100Cには25語登録されています。しかしこれらのフレーズはユーザーが変更できず決められたフレーズのみ認識可能です。

特定話者認識フレーズ(SDフレーズ)はユーザーが任意のフレーズを最大32語登録できますが、フレーズ学習をさせた話者以外の声では著しく認識精度が落ちるため不特定話者の認識には適しません。

本サービスでは、SR-MOD100Cにお客様が指定したフレーズをSIフレーズとして本体に登録、記憶させ納品するものです。これにより最大20フレーズ（10フレーズ×2グループ）のSIフレーズを本体に登録させることができます。

SIフレーズは女性男性問わず日本語であれば話者を問わず認識しますので、様々な機器に組み込んで音声認識による制御、操作が可能になります。

■登録できるSIフレーズについての注意①

音声認識できるSIフレーズは原則として「単語」になります。複数の単語が助詞によって連結している場合には認識精度が悪くなります。

「テレビ を つけて」
└──────────┘ 格助詞

この場合「テレビ」と「つけて」の2つの単語が「を」によって連結されています。このように格助詞によって複数の単語が連結されている場合、後方の単語を異なって発話しても、SR-MOD100Cはなるべく登録されたSIフレーズだとして認識する傾向にあります。例えば同一のグループに

SIフレーズ：「テレビをつけて」
SIフレーズ：「テレビを消して」

という2つのSIフレーズがあった時、正しくフレーズ通り発話すれば認識精度は保たれますが（両者の識別は可能ですが）、「テレビ」だけで発話をやめた場

合や「テレビをどけて」など先頭のフレーズだけ同一でその後のフレーズを変えて場合でも、どちらかの語であると認識することがあります。例えば「テレビをどけて」と発話しても「テレビをつけて」の方として認識することがあります。

これは認識に必要な特異性が求められる単語が不明瞭になった場合、SR-MOD100Cは「認識できないフレーズ」(STS_ERROR)とはせず、2つのうちいずれかのフレーズだとして結果を出力するという事です。これはもともとSR-MOD100Cが認識できるフレーズは単語が基本仕様となっているため、複数の単語から構成される文は前提としていないためです。その点をよくご理解の上、フレーズの選定や機器の設計をして頂く必要があります。



完全一致以外の認識精度についての注意

SR-MOD100CのカスタムSIフレーズでは、登録されたフレーズについて、発話内容が完全一致したもの以外は、似たフレーズを検出します。よって発話内容と認識されたフレーズとは一致した結果にならないことがあります。例えば以下の例を考えてみます。

- ①「でんげんおん」
- ②「でんげんおふ」

この2つのフレーズはいずれも「でんげん」という単語から始まっていますが、発話内容が完全に一致していればSR-MOD100Cは高い認識精度を実現できます。例えば「電源オン」と発話すればSR-MOD100は①と認識して結果を出力します。「電源オフ」と発話すれば②と認識します。

しかし「電源解除」や「電源オープン」など「でんげん」から始まるフレーズで、後方が異なる内容を発話した場合でもSR-MOD100Cは「登録されていないフレーズ」とはせず、①又は②のいずれかとして認識する傾向があります。ただし「電源」だけで発話をやめるとほとんどの場合「登録されていないフレーズ」として結果を出力します。

感覚としては「「でんげん」という単語に何かが続いたフレーズ」としてSR-MOD100Cは判断しているとお考え頂くと分かりやすくなります。この最初の単語以降が多少異なった発話内容であってもSR-MOD100は登録されたフレーズとして認識結果を出力することがありますので、この点についてあらかじめご理解頂けますようお願い致します。

登録するSIフレーズはできる限り短く、単語とすることが推奨されます。また助詞で複数単語を接続する時はなるべく同一の単語が含まれたフレーズを複数作らないようにすると精度が上がります。やむを得ず似たフレーズを登録する必要がある場合にはグループを分けることをおすすめします。

例：

グループ1： 「テレビをつけて」
「鍵を開けて」

グループ2： 「テレビを消して」
「鍵を閉めて」

4つのフレーズを同じグループに入れると似たフレーズは発話品質によって誤認識する可能性が高くなるため、グループに分けて登録し、本機を制御するホスト機器側でグループの制御をするようにします。

■登録できるSIフレーズについての注意②

人間でも電話越しなどで聞き間違えが多い発音はSR-MOD100Cでも認識精度が低下します。例えば「ぶ」と「く」や「でいー」と「びー」など母音が同じ音節で始まる単語の場合注意が必要です。

その後ろに続く音節が類似していなければ多くの場合判別できますが、「ぶれー」と「くれーん」など後方も類似していると認識精度が下がることがあります。また数字の9や10も、「きゅう」「じゅう」(KYUU)(JYUU)と似ているため、先頭の「き」と「じ」を明瞭に発話しないと誤認識してしまうことがあります。

■登録できるSIフレーズについての注意③

1つの単語でも複数の呼び方があるものは統一した呼び方にする必要があります。たとえば「9」は「く」とも「きゅう」とも読みます。SIフレーズ登録ではあくまでも単語をひらがなで表記した時の発音が登録されますので、どちらで認識させるのかを考えて申請して頂く必要があります。「く」で登録した場合「きゅう」と発音しても認識されません。

■登録されたSIフレーズの構成

登録したカスタムSIフレーズはSR-MOD100C本体のSIグループのワードセットインデックス4以降に登録されます。

ワードセットインデックス4は、申請時のSIカスタムフレーズグループ1のトリガーワード、ワードセットインデックス5はグループ1が登録されています。

ワードセットインデックス6は、申請時のSIカスタムフレーズグループ2のトリガーワード、ワードセットインデックス7はグループ2が登録されています。

SIインデックス	登録されている内容
4	グループ1のトリガーワード
5	グループ1のカスタムSIフレーズ

6	グループ2のトリガーワード
7	グループ2のカスタムSIフレーズ

※グループ2は登録申請されていない場合には存在しません。

※SIワードセットインデックス1～3はあらかじめSR-MOD100Cに登録されているSIフレーズです。



トリガーワードとは？

トリガーワードは認識精度を向上させるためのフレーズです。このフレーズを発話して認識を繰り返すことでSR-MOD100Cの認識精度を若干向上させることができます。トリガーワードのフレーズは申請後製品をお届けした際に添付される登録内容案内に記載されています。

EasyVR CommanderにてカスタムSIフレーズ登録がされているSR-MOD100Cを認識させると次のような画面が表示されます。

16	Password	0
--	Sound...	1
1	Wordset	8
2	Wordset	6
3	Wordset	11
4	Grammar	1
5	Grammar	10
6	Grammar	1
7	Grammar	10

上図の四角で囲われている部分がカスタムSIフレーズとなります。新規に登録されたカスタムSIフレーズのグループは"Grammar"という名前がついています。

■登録されたSIフレーズの認識

登録されたカスタムSIフレーズの認識は、あらかじめ登録されているSIフレーズの認識と同様、不特定話者認識プロセスの実行コマンド"CMD_RECOG_SI"(69h)で認識できます。"CMD_RECOG_SI"コマンドに引数として認識させたいフレーズの登録されたワードセットインデックス4以降の値(45h以上)をSR-MOD100Cに送信します。例えばカスタムSIフレーズ、グループ1のフレーズを認識させたい場合には次のコマンドをSR-MOD100Cに送信します。

69h, 46h

"CMD_RECOG_SI"+引数の2バイトを送信後、SR-MOD100Cは音声入力待機状態となります。音声が入

力され認識が完了するとその結果に応じて戻り値が戻ります。登録されたフレーズの場合には"STS_SIMILAR"(73h)が返ります。続いてUARTホストより20hを送信すると、フレーズのインデックスが戻ります。認識ができない場合には"STS_ERROR"(65h)が戻りません。

これらの手順はすべてあらかじめ登録されているS1フレーズの認識プロセスと同じです。

■カスタムSIフレーズ認識の一例

SR-MOD100Cでは下記のようなカスタムSIフレーズを登録した場合、発話内容が一致していればほぼ完全に発話内容を認識することができます。

つけて
けして
あけて
しめて
おふ
おん
きどう
しゅうりょう
さいせい
ていし

ぜんしん
こうたい
みぎせんかい
ひだりせんかい
おおきく
ちいさく
みゅーと
みゅーとかいじょ
めーるじゅしん
めーるをひらく

■免責事項（ご注文前に必ずお読みください）

カスタムSIフレーズ音声認識の精度は発話されたたないようが、登録されたフレーズと一致していれば、かなり高い精度を達成できますが、その音声認識の信頼性は保証されたものではありません。使用については一般的な民生製品での使用の範囲に限定されます。本製品を宇宙、航空、原子力、燃焼制御、運輸、交通、各種安全装置、ライフサポート関連の医療機器等のように、特別な品質・信頼性を要求され、その故障や誤動作が直接人命を脅かしたり、人体に危害を及ぼす恐れのある用途向けには使用できません。

本書1ページに記載しておりますように認識されるカスタムSIフレーズはフレーズ後方が完全に一致していない場合でも、登録されているフレーズとして認識結果を返す場合があります。これはSR-MOD100

Cの性能、特性が発話内容と登録されたフレーズ内容が完全一致したことを判断するものではないことを意味します。発話内容と登録されたフレーズが一致しているか否かではなく、発話内容が似ているか否かを判断する製品であることをご理解ください。

カスタムSIフレーズの登録においては、お客様から申請のあったフレーズを本体に登録後、当方にて実際に発話して認識テストを行い出荷しますが、フレーズの内容によっては認識精度が低い場合があることをご了承ください。

音声認識の精度は、周囲の背景雑音や入力される音声の大きさ、声質、滑舌、方言、イントネーション（強弱のポイント）などによって影響を受けます。例えばくぐもった声や、老人の方の声、はっきりとしていない幼児の声、極端に方言が強い場合などは期待した認識精度が保てないことがあります。

SR-MOD100Cを使用したことにより生じる、明示的及び潜在的な使用危険性や、不確実性について、当方では予見することができません。使用する際には、お客様の責任においてこの製品を正しくお使いいただけますようお願い致します。

本製品を使用することによって生じた、もしくはこれに関連するいかなる直接・間接損害、懲罰的損害、その他データの破損や消失等を含むいかなる損害、損失についても、当方では一切責任を負いかねます。あらかじめご理解とご了承頂きますようお願い致します。また認識精度が高くないことが原因で発生したいかなる問題、損失についても当方で一切その責任を負いかねます。

本製品を使用した製品等を製造される場合には様々なフェイルセーフ機能（安全設計）を施して頂き、十分に機器のテストをした上で運用されますようお願い致します。また、データの損失や予期しない事態に備え、データのバックアップを行って頂きますようお願い致します。

マイクロテクニカ



〒158-0094 東京都世田谷区玉川1-3-10

TEL: 03-3700-3535 FAX: 03-3700-3548

(C)2017 Microtechnica All rights reserved