

センサリ社音声技術説明

センサリ社(Sensory,Inc.)が提供している音声技術を説明いたします。

IC 製品 : NLP-5x、RSC-4x、SC-6x とソフトウェア製品 : FluentSoft™、BlueGenie™などに関連した技術となります。

◆ 音声・音響技術

センサリ社の IC 製品は、先進の HMM(Hidden Markov Modeling)やニューラルネットワーク音声認識、音声合成、録音再生などの音声技術と IC テクノロジを使った製品、組み込みソフトウェア製品で、最先端の機能を非常に低価格で提供します。

■ ■ 音声認識技術 ■ ■

センサリ社は独自の HMM とニューラルネットワーク技術を開発しこの技術を採用することで高価な DSP や容量の大きな RAM を一切使わずに、信頼性の高い音声認識を提供しています。

特に独自のニューラルネットワーク技術は特定の話者による単語やフレーズを識別することのみならずセキュリティを向上させるために使用することもできます。センサリ社の 20 年以上に渡るスピーチとオーディオテクノロジーの経験を使用することでセンサリ社の IC 製品は高品質なオーディオ出力を実現でき音質を劣化させることなく音声や効果音を再現することができます。

またセンサリ社のソフトウェア製品は組み込み実装を目的とした小型ソフトウェアエンジンとなっています。長年 IC 製品で培ってきた HMM&ニューラルネットワークの音声認識技術を応用し実現しています。

● フレキシブル・グラマー <Flexible Grammar>

【NEW】

対応 : NLP-5x、FluentSoft™

センサリ社 NLP-5xの自然発音インターフェースはユーザーの命令発音に準じて認識することのできる技術です。認識する順番に捉われない方法で認識し単語が足りない場合には音声合成技術を利用して足りない単語を確認するなどができます。この画期的なフレキシブル・グラマー技術は複数の単語を一フレーズとしては認識することができ、且つ命令語を順番に発音する必要もないためもっとも自然な認識を可能としました。

使用例) オープンなどの設定

例 1) ユーザー:「オープン、120°Cで、1 時間 45 分」

オープン:「設定しました」

例 2) ユーザー:「オープン、120°C」……時間情報が言われなかった。

オープン:「時間を指定してください、何時間何分ですか？」

ユーザー:「1 時間 45 分」

● フレーズスポッティング <Phrase Spotting>

【NEW】

対応：NLP-5x、FluentSoft™

完全なハンズフリー操作。

最大 30 語の命令のキーワード・スポッティングやフレーズスポッティングが騒音下でも可能です。会話中のトリガーワードを認識し音声認識操作に移行することができ完全なハンズフリー操作が実現できます。

● 不特定話者音声認識 <Speaker Independent WithT2SI>

対応：NLP-5x、RSC-4x、FluentSoft™、BlueGenie™

言葉の登録は必要無し。

不特定話者音声認識' Speaker Independent'とは誰の音声でも認識するという音声認識技術です。予め決まった言葉の辞書を事前に保存しこの言葉のみを認識させるというものです。この音声認識技術は汎用品に利用されます。センサリ社の不特定話者音声認識は独自の HMM (Hidden Markov Modeling)&ニューラルネットワーク技術を使っています。この技術はRSC4xシリーズとFluentSpeech™製品で利用されており、高速・高性能な認識を実現しています。不特定話者音声認識ではコンティニュアスリスニング、ワードスポッティングをサポートしています。

● 特定話者音声認識 <Speaker Dependent>

対応：NLP-5c、RSC-4x

辞書構成、言語、発音に影響されません。

特定話者音声認識' Speaker Dependent'とは使用者が音声を登録して始めて使用できるようになる音声認識技術です。予め決まった言葉の辞書は保存されていないのでユーザーが任意に言葉を登録することができます。この音声認識は個人または数人で利用するなどの商品に利用されます。この技術は音声のみならず登録できる「音」であれば認識できるもの応用として効果音や動物の鳴き声など登録し認識させることができますセンサリ社の音声認識には主に独自のニューラルネットワーク技術を使っています。

RSC4xシリーズでは従来のニューラルネットワーク技術に加え HMM (Hidden Markov Modeling) 技術を利用し、より高速高認識率を実現しています。特定話者音声認識ではコンティニュアスリスニング、ワードスポッティングをサポートしています。

RSC4xシリーズは'不特定+特定話者音声認識(T2SISD)'をサポートしています。不特定話者の登録単語と特定話者の登録単語を同時に認識することができます。

● 音声認証(パスワード認証) <Speaker Verification>

対応：NLP-5x、RSC-4x

音声のバイオセキュリティ。

音声認証(パスワード認証)'Speaker Verification'は音声を登録音声と比較し認証する技術です。話者認証として利用できます。センサリ社の音声認証は'パスワード認証'というもので特定のパスワードを認証するものです。特定話者音声認識と近い技術ですが特定話者音声認識は一音声入力単語を複数の単語と比較し、結果として一致したものを戻すというのですがパスワード認証は一對一で比較しどの程度一致しているかを確認し結果を戻すというものです。

■ ■ オーディオ技術 ■ ■

音楽や音声出力に対応したさまざまな技術を提供しています。

● テキスト音声合成 <Text-To-Speech> 【NEW】

対応 : NLP-5x、FluentSoft™、BlueGenie™

テキストから音声合成を再生するソフトウェアエンジンです。

本技術はメールなどの文字を読み上げるなどに利用するための技術です。現在はPCおよびPDAプラットフォームに実装できるソフトウェアのみでセンサリ社プロセッサおよび他社プロセッサ実装には対応していません。

また本テキスト音声合成(TTS)技術はボイス・モーフィングに対応しており再生する音声を音程などを変化させ、男性、女性、キャラクターなどで再生することができます。

* NLP-5xは US 英語のみ対応(スペイン語、フランス語、ドイツ語、中国語、日本語は開発中)。

* FluentSoft™は US 英語、スペイン語、フランス語対応(中国語、日本語は開発予定)。

* BlueGenie™は US 英語対応(ドイツ語、中国語は開発中)

アプリケーション: 英語発話玩具、電子辞書、など。

● MP3 デコーダー <Stereo MP3 Decoder>

対応 : NLP-5x

高音質ステレオ MP3 デコーダー。

全てのビットレート対応、5 バンドのイコライザー機能。

アプリケーション: 音楽プレーヤーつき玩具、携帯音楽プレーヤー、読み聞かせ絵本、など。

● 音楽合成 <Mono/Stereo Music>

対応 : NLP-5x、RSC-4x、SC-6x

音楽合成' Music Synthesis'は NLP-5x ではサンプルレート 32K サンプル/秒で最大 24 ステレオボイスを生成できます。RSC-4x ではサンプルレート 8K サンプル/秒で最大 8 モノラルボイスを生成できます。音楽は IC のステレオ DAC 出力かモノラル PWM から SX,PCMADPWM 音声、効果音などを重ね合わせて再生できます。音楽合成技術は内蔵メモリー、シリアル Flash などの外部メモリーに保存された MIDI 音楽を再生できます。この方法はメモリー効率の良い方法となります。

音楽合成は楽器音のオーディオデータベースが必要です。センサリ社では General MIDI メロディー楽

器、パーカッション楽器などの標準的な楽器データベースと提供しています。これらの楽器データベースは通常外部メモリーに保管されます。

アプリケーション： 楽器玩具、音楽再生玩具、簡易カラオケ玩具、など。

● 音声合成 <Speech Synthesis>

対応：NLP-5x、RSC-4x、SC-6x、FluentSoft™

音声プロンプトに最適。

音声合成' Speech Synthesis'は圧縮音声再生する技術です。

センサリ社では複数の圧縮方式を採用しています。これら技術は従来の PCM、ADPCM、MELP、CELP などの組み合わせたセンサリ社独自方式の'SX'音声合成などがあります。圧縮音声の再生レートは 1.0Kbps の低圧縮から 68Kbps の高音質のまでの音声合成を提供しています。音声合成データへの変換は専用ソフトを利用すれば簡単に行えます。これら音声合成の原音は WAV ファイルを利用しています。

アプリケーション： お喋り人形、お話絵本、など。

● 録音再生 <Record and Playback>

対応：NLP-5x、RSC-4x

デジタル録音、メッセージの録音と再生。

RSC-4xと NLP-5xは 30Kbps 以下で録音、再生することができます。この録音データは外部メモリーに保存することができます。これらはグリーティング、電話、留守電、ボイスチェンジャー、録音デバイスに利用できます。

音質に合わせて幾つかの圧縮レベルを設定できます。また自動無音圧縮を行いメモリー効率で保存することができます。RSC-4x は 8K サンプル/秒。NLP-5xはより高音質で録音再生ができます。

アプリケーション： 簡易 IC レコーダー、録音再生玩具、ボイスチェンジャー玩具、など。

■■ インターラクティブ・ロボティック技術 <Interactive/Robotic> ■■

LCD(液晶)、モーター制御、データ通信などさまざまな制御に対応した技術を提供しています。

● LCD 制御 <LCD Control>

対応：NLP-5x

LCD 制御回路。104 個のアイコンまたはピクセルに対応。大規模アレードライバに対応した SPI を準備。

● モーター制御 <Motor Control>

対応：NLP-5x

モーター制御回路。最大3バイ・ディレクショナル(正逆3方向)モーター制御可能。

● サイレント・ソニックネット技術 <Silent Sonic Net>

【NEW】

対応：NLP-5x

サイレント・ソニックネットは NLP-5x で 14KHz または 18KHz のエンコード音でデータのバースト通信ができます。結果としてこれら高周波音は無音（耳に聞こえない）となり、通常のアプリケーション中に短くバーストデータ転送できますサイレント・ソニックネットは SX 音声合成または T2SI の音声認識中にデータ通信が動作します。NLP-5x とマイク、スピーカー実装の製品では追加部品を必要とせず製品間のデータ通信などができます。

アプリケーション： 対話機能対応の人形（ぬいぐるみ）、DVD 連動玩具、時計同期して喋る人形など信号に同期して動作させたいなどの製品。

● ソニックネット技術 <Sonic Net>

対応：NLP-5x、RSc-4x

「ソニックネット(SonicNet)」技術は音声認識マイコン'RSC4x'にて可聴域音によりデータ通信する技術です。音声合成に組み込むなど、または単独に通信するなどが可能です。この技術により赤外通信などを必要とせず音声認識製品で利用されている通常のマイクとスピーカーにて機器同士の通信が可能です。

アプリケーション： 対話機能対応の人形（ぬいぐるみ）、DVD 連動玩具など。

● 通信インターフェース <System Communications>

対応：NLP-5x

USB1.1、SPI、UART-Lite、I2S、IR インターフェースを準備。

● リアルタイムリップシンク技術 <Real Time LipSync>

対応：NLP-5x、RSC-4x

「リアルタイムリップシンク(Real Time LipSync)」技術は音声解析技術を応用しメカニカルな口を制御する技術です。音声認識の解析技術を応用しているため音声のみに反応するように調整されていて音楽の中のボーカル音声のみに口が同期して動くなどの商品に利用します。メカニカルな口への制御はモーター制御用の PWM 信号で行います。

*音声帯域に反応するように調整してありますが音や楽器によっては反応してしまうことがあります。

アプリケーション： 歌に合わせて口が動くオーディオプレーヤー人形、など。

● リップシンク技術 <LipSync>

対応：NLP-5x、RSC-4x

「リップシンク(LipSync)」技術は音声合成に同期して口メカを制御するものです。

アプリケーション： 口が動くお喋り人形、など。

● ビート予測ビート検出技術 <Beat Predict/Beat Detect>

対応：NLP-5x、RSC-4x

「ビート予測ビート検出(BeatPredict/BeatDetect)」技術は音楽のビート(テンポ)を検出する最新の技術です。ビート予測・検出は音楽(歌)などに含まれる特定の音の一定のテンポを検出、解析しビート信号を出力するものです。

アプリケーション：音楽に合わせて動くダンス人形、ライト点滅、など。

● ピーク検出技術 <Peak Detection>

対応：NLP-5x、RSC-4x

音の振幅ピークを検出する技術です。

アプリケーション：音楽に合わせて動くダンス人形、ライト点滅、など。

● ピッチ検出技術 <Pitch Detection>

対応：NLP-5x、RSC-4x

音のピッチを検出する技術です。

アプリケーション：歌った音声の音程を解析するゲームなど。

● シングバック技術 <Sing Back>

対応：NLP-5x、RSC-4x

「歌まね(Singback)」技術はトークバック技術とピッチ検出技術を組み合わせユーザーの「歌」などの音声(音調、音程)を記憶し、合成音声にて再生(繰り返す)する技術です。音声や鼻歌の音程を解析して音程情報を出力することができるため、音声合成を組み合わせ音声を楽器音などにて出力するなどが可能です。

アプリケーション：歌に合わせて同じ音程で歌い返してくれる人形、など。

● トークバック技術 <Talk Back>

対応：NLP-5x、RSC-4x

「トークバック(Talkback)」技術はユーザーの音声などの音声(音調、音程)に反応し 合成音声にて特定の音声を再生する技術です。

アプリケーション：喋り掛けると同じような音声(音程)で喋り返してくれる人形、など。

● 音源検出技術 <Sound Sourcing>

対応：NLP-5x、RSc-4x

「音源検出(Sound Sourcing)」技術は音源(音の発生源)を検出する最新の技術です。「音源検出」は音の方向は 2 つのマイクで拾い、音量差を計算しマイクマウントをモーターにフィードバック制御するこ

とにより正確な音源方向にマイクを向けるなどができます。

アプリケーション：音に向かってくるロボット、音に向かって首を振る人形、など。

● ナチュラルタイムセット技術 <Natural Time Set>

対応：NLP-5x、RSC-4x

「ナチュラルタイムセット(NaturalTimeSet)」技術は時刻を設定する技術です。通常のマイコン音声認識では時間、分を分けて音声認識させないとならなかったのですがこの技術により「12 時 34 分 PM」などと連続音声でも認識できるようになりました。このことにより時刻設定の煩わしさが軽減されました。(現在は英語のみ)

アプリケーション：時計、オーディオ機器、家電機器などの時刻、アラーム、動作時刻設定。

● ナチュラルデートセット技術 <Natural Date Set>

対応：NLP-5x

「ナチュラルデートセット(NaturalDateSet)」技術は日付を設定する技術です。NLP-5x でデジタル時計などの日付を自然な言い方で設定できます。

アプリケーション：時計、オーディオ機器、家電機器などの時刻、アラーム、動作時刻設定。

● オーディオウエイクアップ技術 <Power Audio Wakeup>

対応：RSC-4x

「オーディオウエイクアップ(AudioWakeUp)」技術は省電力モードから音によりICをウエイクアップ(起動)する技術です。これにより常に音確認をする動作をせずウエイクアップ音の入力によって認識を開始することができ特にバッテリー駆動製品など電力消費を抑えるなどの改善ができます。

■ ■ ブルートゥース向け音声認識技術 <Voice Recognition for Bluetooth Products> ■ ■

● BlueGenie™音声インターフェース <Blue Genie Voice Interface>

対応：BlueGenie™

Bluetooth 製品(ヘッドセット、音楽プレイヤー、ハンズフリー向け音声認識とテキスト音声合成 BlueGenie™(ブルージニー)音声インターフェースソフトウェアは英国 CSR 社 BC-5MM の KalimbaDSP で動作するソフトウェアです。Bluetooth 製品メーカーには不特定話者音声認識、音声合成、テキスト音声合成を利用することでマニュアル無し、完全に音声だけで操作できる Bluetooth 製品を作ることのできるソフトウェアです。テキスト音声合成は着信時の発信者番号の読み上げや SMS の読み上げなどに利用可能です。

アプリケーション：Bluetooth ヘッドセット、Bluetooth ステレオヘッドセット、Bluetooth ハンズフリー・カーキットなど。

■■アニメーション(Lip 同期) <Animated Speech (Lip Sync)> ■■

インタラクティブ製品用として開発されたフェーシャル・アニメーション用ソフトウェアエンジンです。音声に合わせて、唇を同期させて喋るアニメーション技術で、多くの発音データを解析し作成されたデータにて同期動作しますので非常に正確な発音を実現できます。Fluent Speech™, Fluent TTS™ との組み合わせが可能です。また WAVE ファイルに同期再生することもできます。お客様から提供されたサンプルテキスト(英文)をアニメーション&TTS 加工いたします(4 センテンスまで)。

* 本技術は開発中のため、個別にお問い合わせください。

詳細は下記問い合わせ先までご連絡ください。

センサリージャパン株式会社

〒241-0004 横浜市旭区中白根 1-26-28

TEL:045-958-1730/FAX:045-958-1731

Email: sales@sensory.co.jp

2010/3/30