IoT eラーニング

IoTとインターフェース②

（IoTにおける音声入力・音声出力・音声解析）

国立大学法人　琉球大学

目次

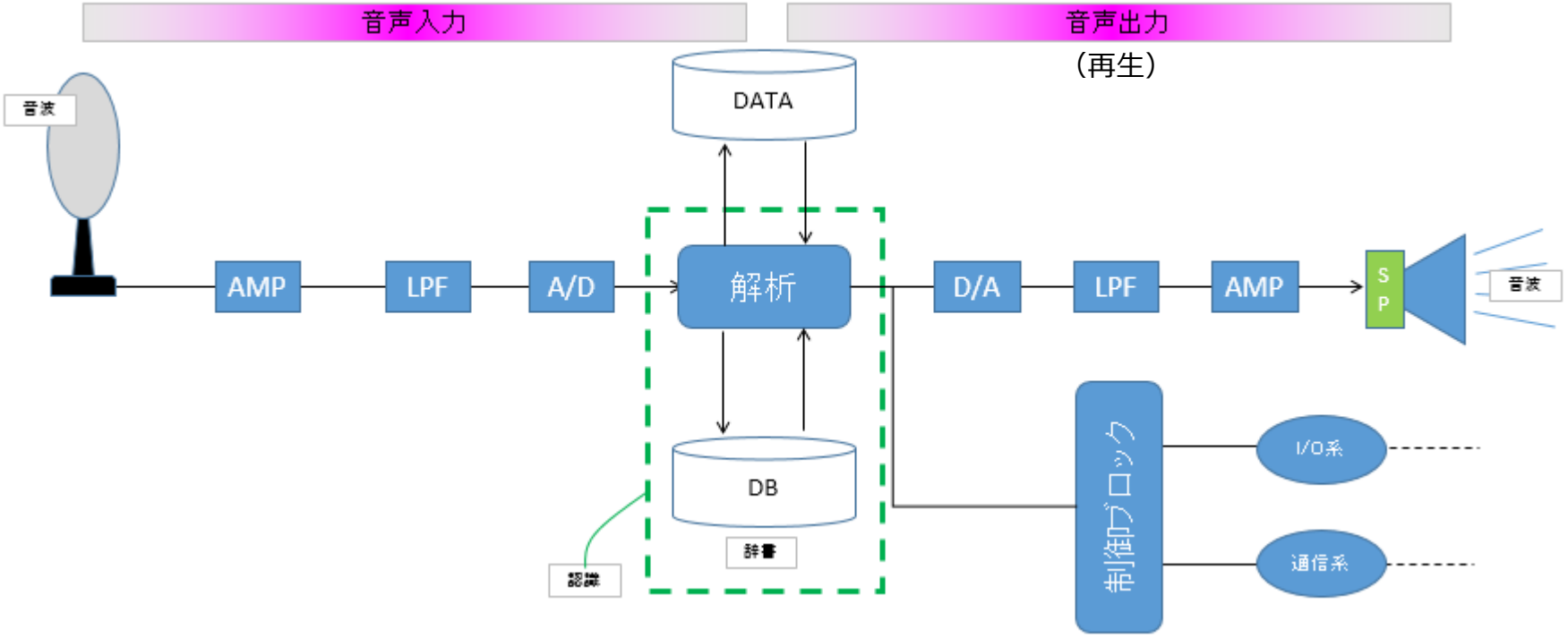
- 音声入力から出力まで
  - 概要
- 音声入力
  - 音声入力の仕組み
- 音声出力
  - 音声出力の仕組み
- 音声解析
  - 人の音声
  - 波を解析する（FFT）
  - 窓関数
  - ケプストラム
  - 音声認識
- IoT向け音声入出力の実際
  - 音声合成LSI
  - 音声録音再生IC
  - 音声認識モジュール
  - 音声認識を利用したシステム

音声
入力から出力まで

音声入力から出力まで

- **概要**

  音声でWEB検索や家電品のコントロールなどのDEMOを目にする機会が増えています。
  その全体像はどのようになっているのでしょうか。
  図は、音声入力から音声出力までを簡単に描いたものです。
  IoTに限らず、図のような処理ブロックで、音声による制御システムが構築されています。



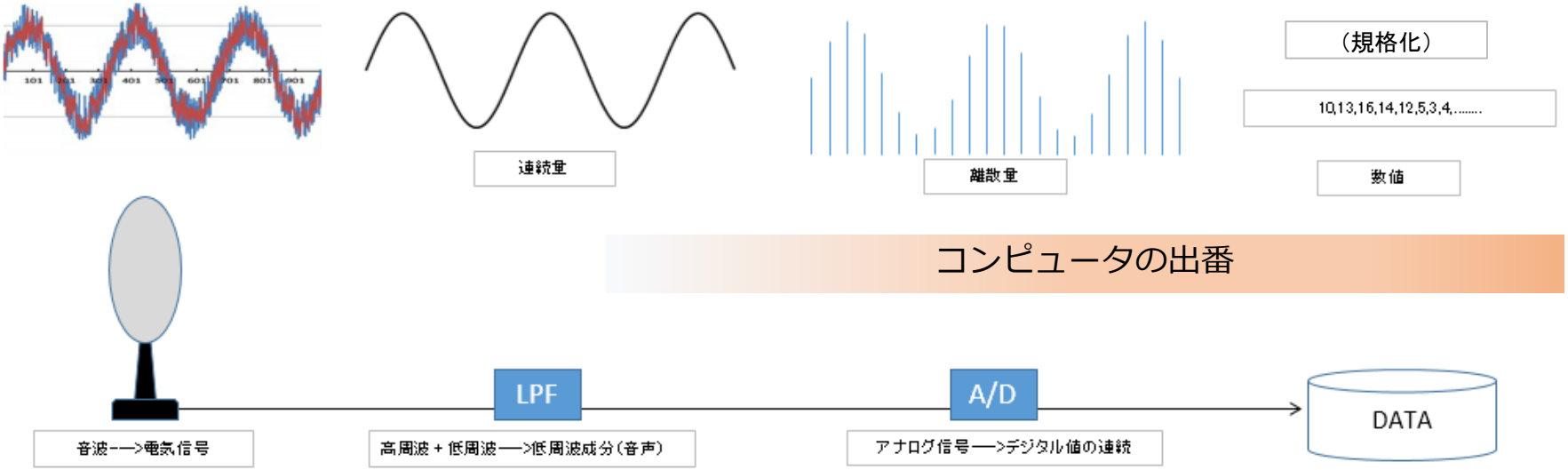
音声入力

音声入力

- **音声入力の仕組み**

  音は空気の波です。これをIoTで利用するために、デジタルの値に変換します。
  生の音をマイクで拾い、耳で聞こえる音の成分をLPF（ローパスフィルタ）で取り出し、A/D変換でデジタル値に変換します。変換結果は数値となってコンピュータで処理ができる値になります。このデジタル値を、取得周期※を含めてファイル保存することで、デジタル録音ができます。高速処理が可能なCPUが開発されて、電気回路で実現していたフィルタなども、デジタルフィルタとして、プロセッサで処理することができるようになっています。
  PC用マイクでUSBコネクタ接続するものは、音声がデジタルデータとなって送信されています。保存するデジタル値は、規格化（例：最小または基準データを１として表現）されたものにする場合もあります。

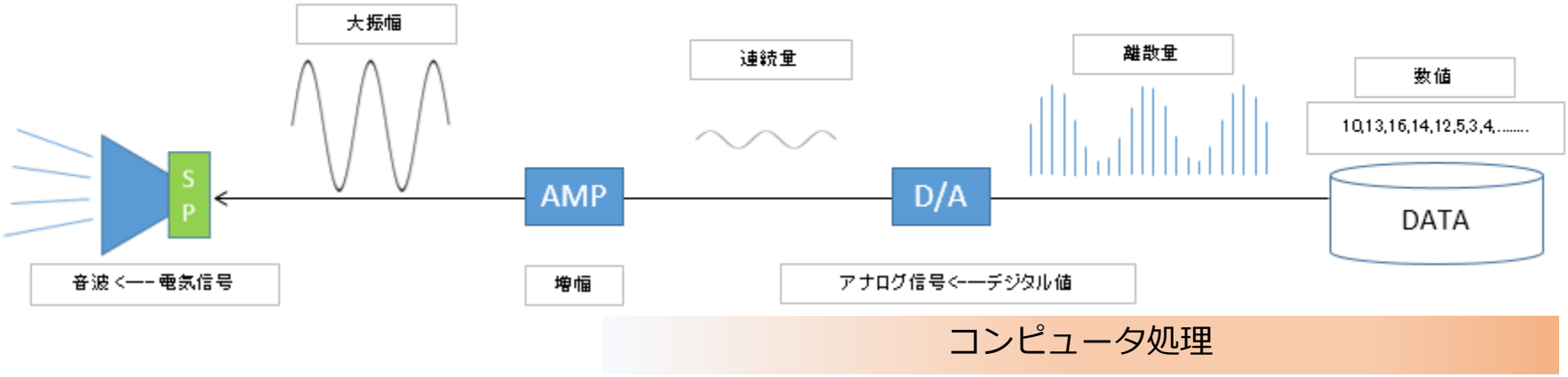
  ※元の周期の1/2以下→ナイキストの定理



音声出力

音声出力

- **音声出力の仕組み**
  デジタルデータとして保存された音のデータは、数値情報なので、それを記録時の時間間隔で、D/A変換（デジタル→アナログ）すると、連続した信号が得られます。この連続信号は、デジタルデータとして保存した際のスケールファクタにより、小さくなっているので増幅やオフセット補正などを行い、スピーカなどから音波として出力されます。

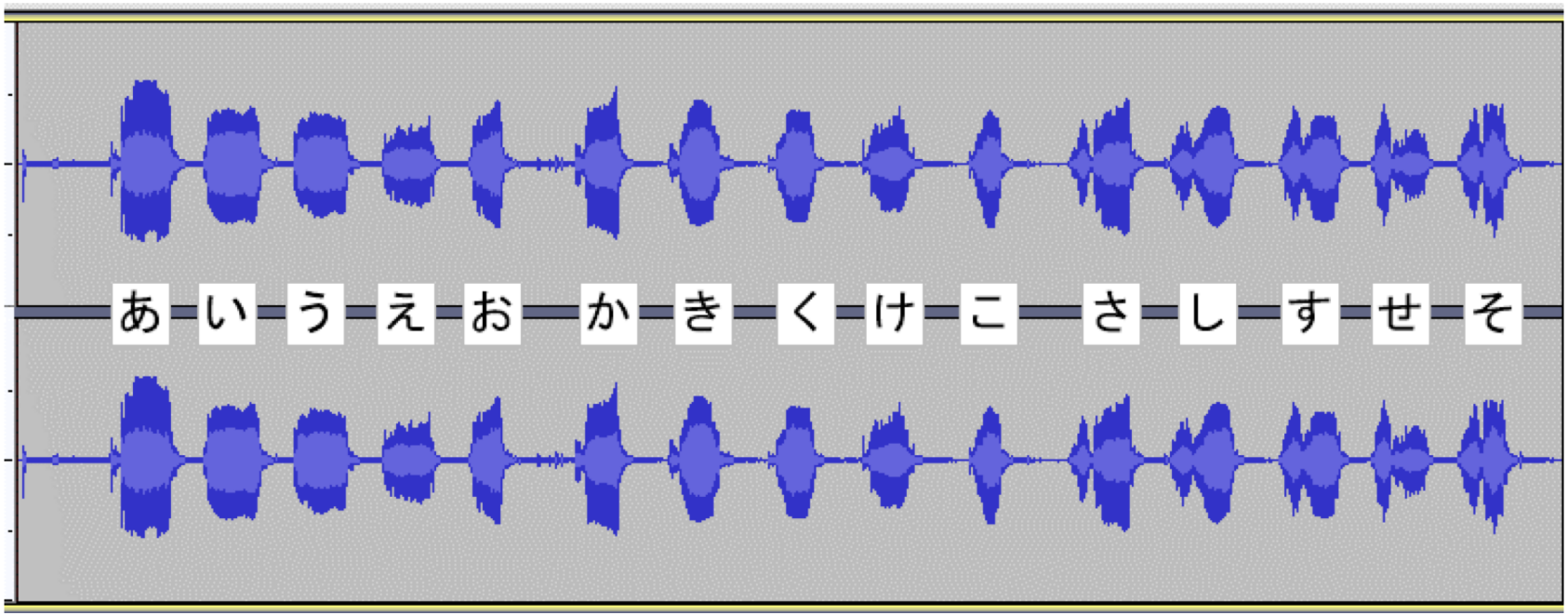


音声解析

音声解析

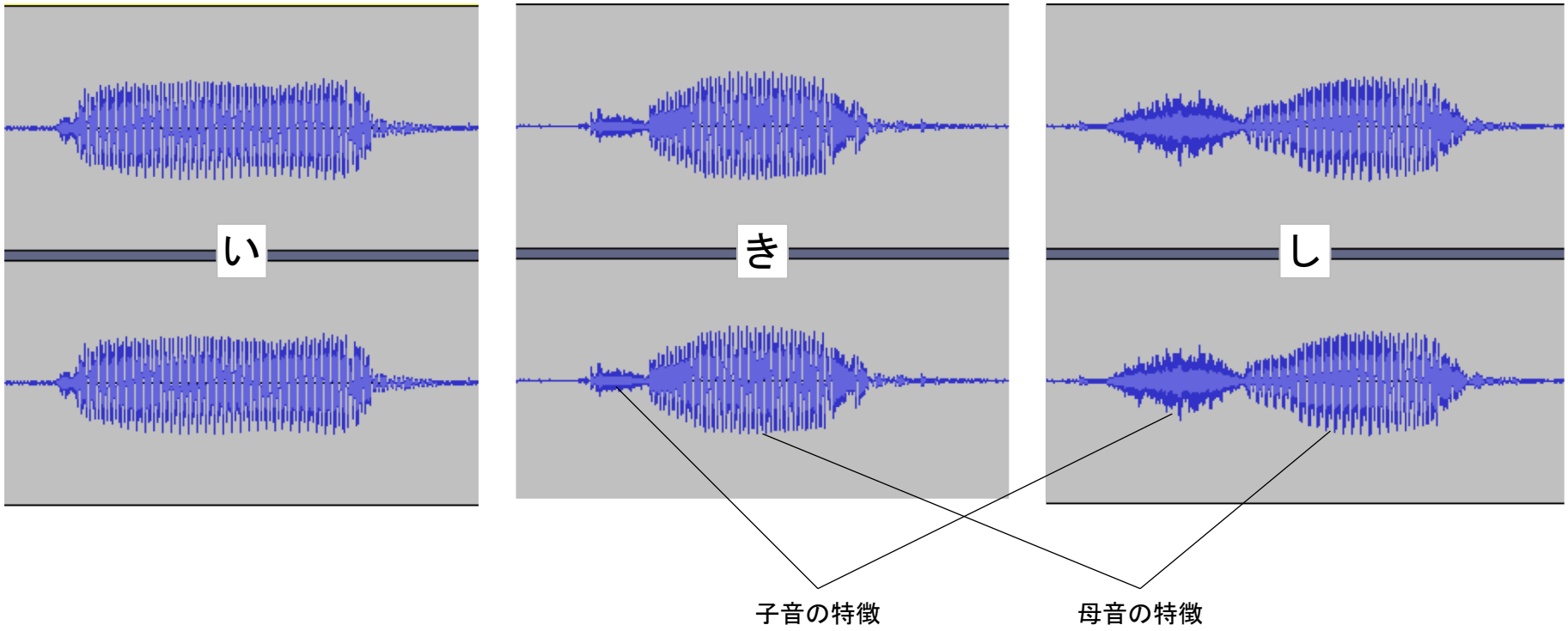
- **人の音声**

図は、私が「あいうえお・・・さしすせそ」を読み上げたときの音声を時間と音波の強さ（電圧）として、表示したものです。音声には母音と子音がありますが、その特徴が見えています。　　※左右のチャンネルを別々に表示していますが、その差はありません。



- **人の音声**

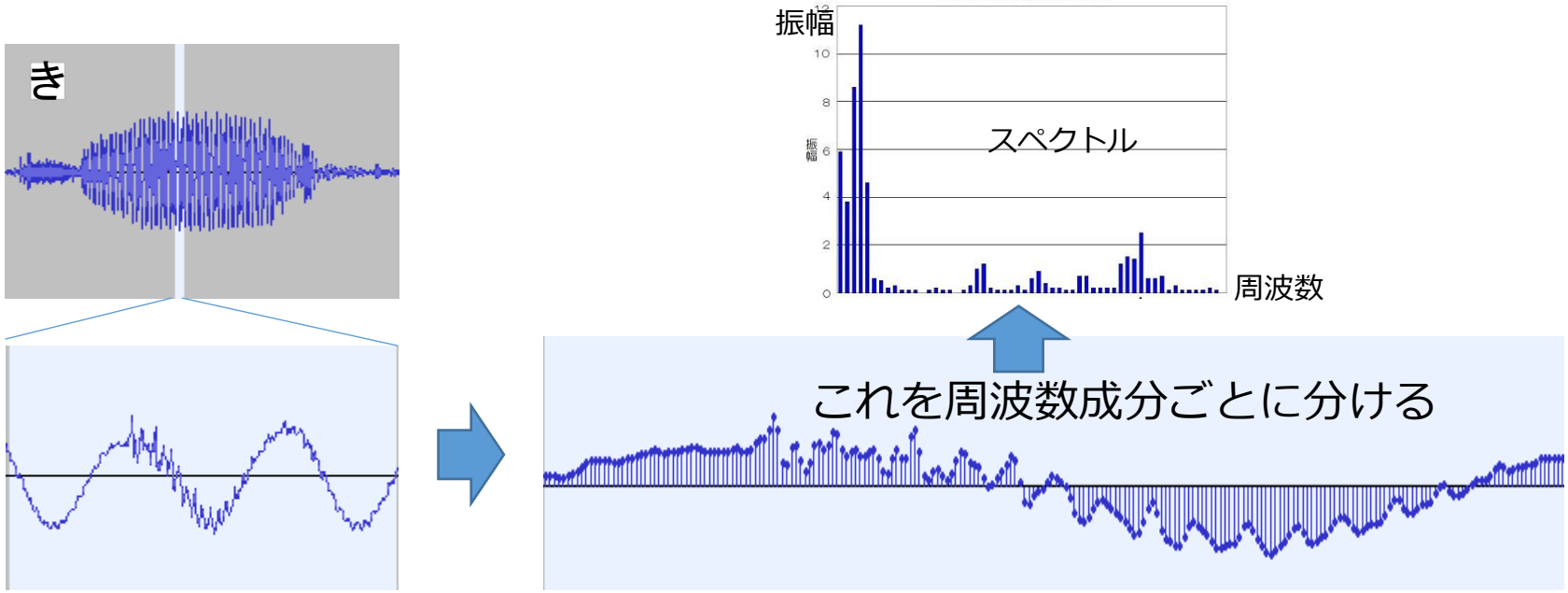
【い】の母音をもつ音の波形を拡大してみます。
【き】【し】は、子音の後に、確かに母音の【い】が付いている事が分かります。



音声解析

- **波を解析する（FFT）**

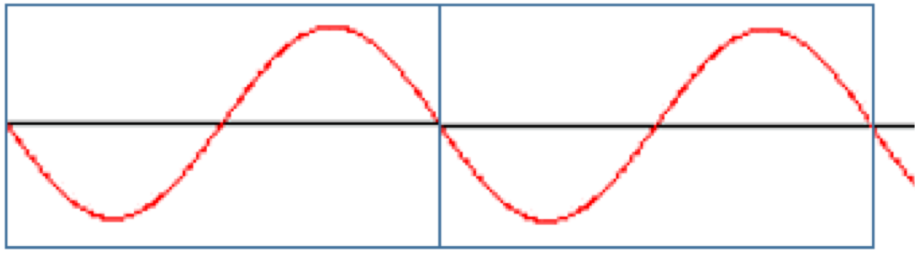
  音声に含まれている、波の様子を解析することが音声解析です。
  波の解析となれば、それは、周波数と振幅を調べることです。
  音の場合、周波数は時間に対応し、マイクロフォンで取得する音の振幅は電圧に対応します。
  音声には色々な周波数の波が含まれています。信号がどんな周波数成分（スペクトル）を持っているかを調べることが音声解析になります。
  これを周波数解析やスペクトル解析と呼んでいます。
  周波数解析には、高速フーリエ変換（FFT）のアルゴリズムが利用できるので、デジタルデータとして記録したものが使えます。このデジタルデータは一定のサンプリング周期で取得したデータで、時間間隔を持つため、離散フーリエ変換ともいわれます。



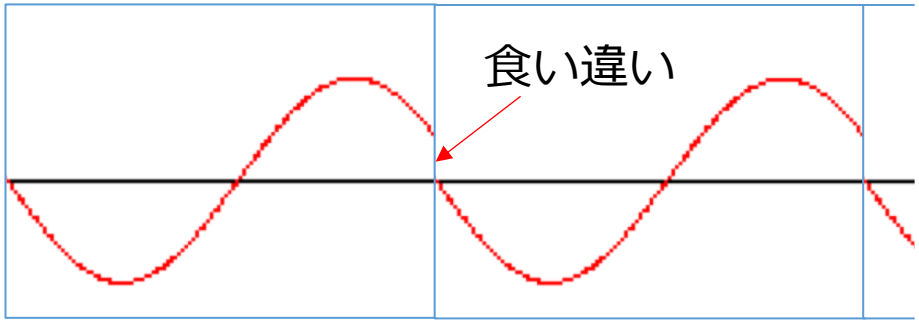
- **窓関数**

  実際には、データ全てについてフーリエ変換を行うのではなく、一定期間のデータを採用して、それをつないで処理します。その時、単純につなぎ合わせると、図Aの様に巧い区間で区切ることができる場合は良いのですが、実際には対象データの周期など分かっていないので、つなぎ目が食い違ってしまいます。
  そこで、図Cのような関数をかけてやると、両端が滑らかに 0 になり、つなぎ目が巧くつながります。このような関数を窓関数と云い、目的に応じて関数を選択します。
  このようにして、フーリエ変換を行い、音声の特徴を取り出します。

図A



図B



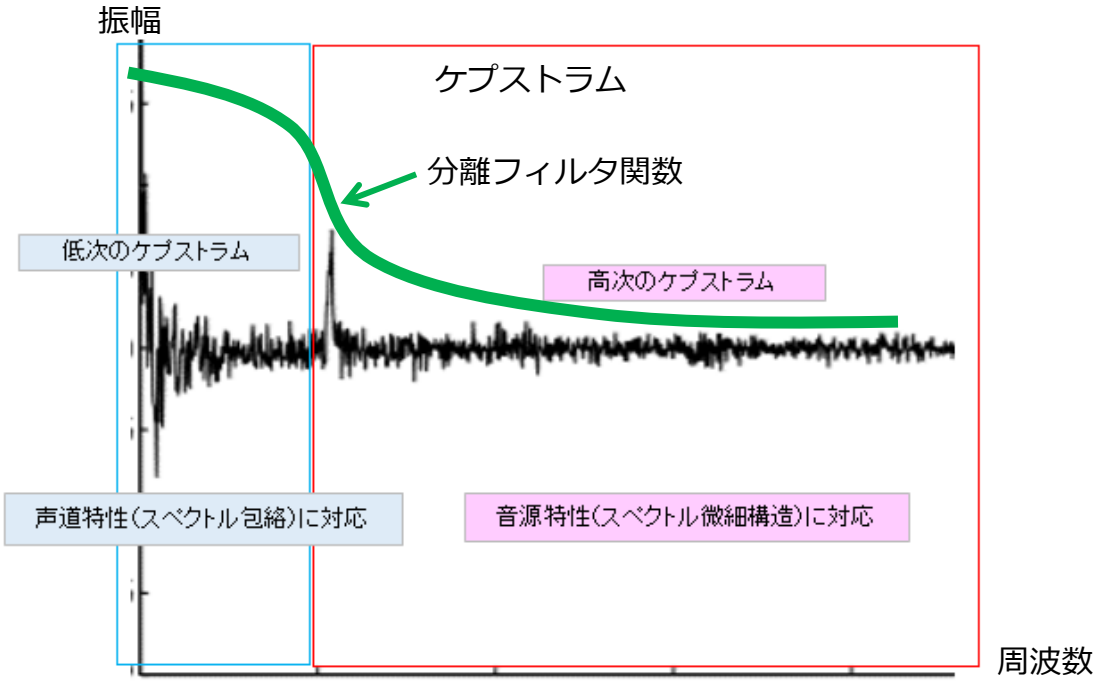
窓関数の例

図C

音声解析

- **ケプストラム**

  音声スペクトルを対数変換してさらにフーリエ変換すると、音声の特徴量である値【ケプストラム】と呼ばれる値を取り出すことができる。
  ケプストラムを使うとスペクトルの細かな変動とスペクトルのなだらかな変動【スペクトル包絡】を分離できます。
  【スペクトル包絡】は人間の声道の特性を表しているため音声認識や音声合成では重要な情報として取り扱われます。低次と高次のケプストラムは、フィルタ関数で分離することができます。



音声解析

- **音声認識**

  音声を解析してテキストとして出力する事が音声認識です。
  音声認識は，コンピュータ上の「音響モデル」と，入力された音声信号（波形）のマッチング（照合）によって行います。
  「音響モデル」には平均的な発音データを基に作られた，音声の単語辞書が使われます。
  マッチングには【HMM(Hidden Markov Model)】（隠れマルコフモデル）理論が用いられることが多い。
  辞書マッチングは10～20ミリ秒の間隔で，単語の先頭から順次処理されます。

  例：「おんせい」という言葉を認識する場合
  ①．最初の「お」という言葉を認識した時点で、マッチングの候補は単語辞書に載っている「お」から始まる言葉に絞られます。
  ②．次に「おん」を認識した時点で候補は，音楽，温泉，音声・・・に，
  ③．さらに「おんせ」を認識した時点では，温泉，音声・・・に，といった具合です。
  ④．どんどん候補が絞り込まれて，最終的に最もマッチングする単語が結果として出力（認識）されます。

  ※ 辞書にない単語は未知語として扱われて，認識されません。
  あらかじめ用途やCPU処理能力，メモリー容量を考慮してチューニングした辞書を使用するようにします。

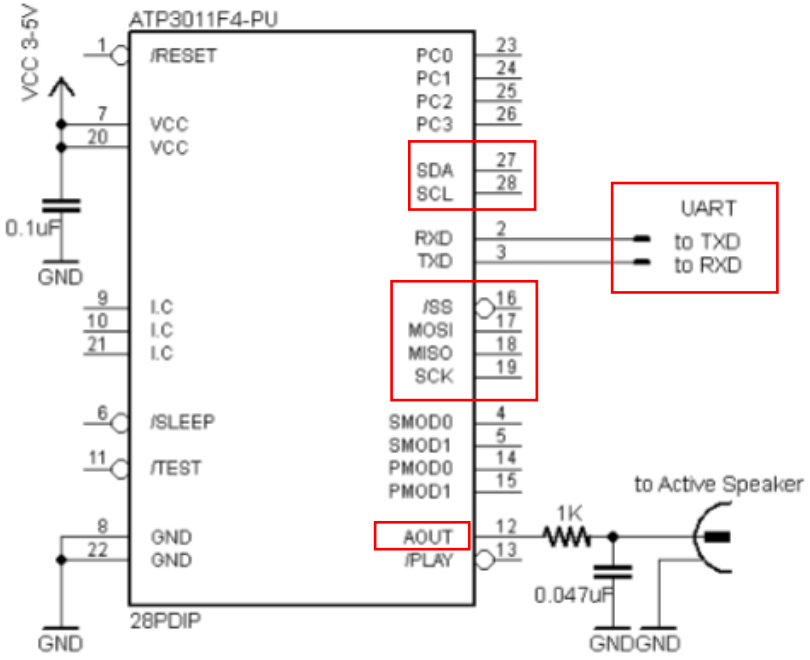
IoT向け
音声入出力の実際

IoT向け音声入出力の実際

- **音声合成LSI**

  録音せずに音声を発するLSIがあります。このLSIはマイクロコンピュータそのものです。
  内部フラッシュメモリに音声合成ミドルウエアをファームウェアとして書込んだものです。
  UARTポートからシリアル通信で、発音させたい言葉の【ローマ字読み】を文字列として送り込めば、AOUT端子から合成音性が出力されます。（規則音声合成）
  IoTとしては、マイコンからシリアル通信で発音内容が制御できるので、録音機器、録音の手間が無く便利です。音声も男性、女性、ロボットなどが選べます。

内部はATmega328P



- **音声録音再生IC**

APR9600 RE-Recording Voive IC

Single-Chip Voice Recording & Playback Device

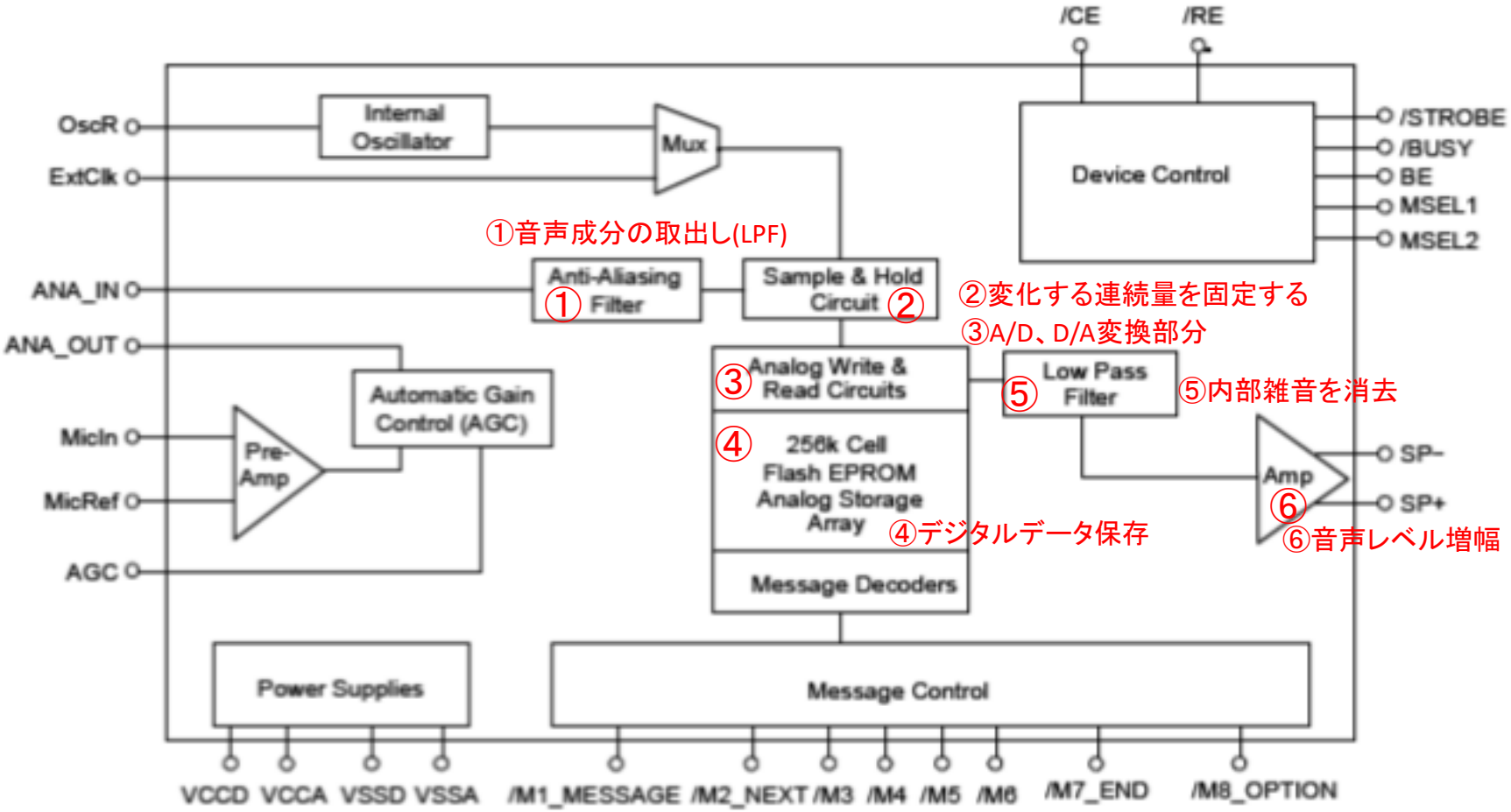
60-Second Duration

1 Features :

- Single-chip, high-quality voice recording & playback solution
  - No external ICs required
  - Minimum external components
- Non-volatile Flash memory technology
  - No battery backup required
- User-Selectable messaging options
  - Random access of multiple fixed-duration messages
  - Sequential access of multiple variable-duration messages
- User-friendly, easy-to-use operation
  - Programming & development systems not required

IoT向け音声入出力の実際

- **音声録音再生IC**
  デジタル録音・再生の機能を1チップに収めたICの（Aplus：APR9600）のブロックチャートを示します。内部の仕組みは、先に示した音声入力と音声出力の処理ステップに従っている事が分かります。



IoT向け音声入出力の実際

- **音声認識モジュール**

  音声認識をIoTで容易に利用できるように、組み込みモジュールとして開発された製品も容易に入手できます。小さな基板に纏められ、音声による制御用出力が数点取り出せます。使用者がマイクを通して録音したフレーズで制御を行います。

  写真左：汎用のI/Oを制御するもの

  写真右：シリアル通信で特定の電文を送信するもの（不特定の話者にも対応）

IoT向け音声入出力の実際

- **音声認識を利用したシステム**
  音声認識を実際に利用したシステムが身の回りに多くなってきました。

AIスピーカ


関西弁ルンバ


家電コントローラ

ハイブリッド字幕放送


物流ピッキング

◇次回のテーマは・・・

IoTにおけるモーション検知・解析