

定番の音声合成方式  
「テキスト読み上げ」のメカニズム

高橋 隆雄, 山崎 信英

ここでは、音声合成の基本方式である録音編集方式とテキスト読み上げ方式(TTS)の仕組みを解説します。

(編集部)



## 二つの音声合成方式

かつて、コンピュータを「しゃべらせる」ことは一種の憧れでした。パソコンがまだ8ビットで、せいぜいビーブ音程度しか出ないころであっても、パソコンにしゃべらせようという試みはさまざまな方法で行われてきました。近頃のパソコンは、音声や音楽、動画をも再生できるような処理速度を持ったため、パソコンでの音声合成はそれほど難しいものではなくなりました。

最近になって音声合成の機能はパソコンの枠を超え、今やマイコンでも十分に扱えるようになってきています。これは図1に示すように、マイコンの性能が上がったことと、コンパクトな音声合成処理アルゴリズムが普及してきたことによります。

## ● 録音編集方式とテキスト読み上げ方式がある

音声合成の方式にはさまざまなものがあります。各社からいろいろな音声合成システムが出されていますが、合成

品質、動作環境も異なります。自動車がエンジンの種類や駆動方式、ボディ形状などのさまざまな観点から分類されるように、音声合成技術も、一元的に分類することはできません。実際の音声合成システムは、複数の技術が絡み合って構成されているので、着目する視点によって異なる分類方法があります。そこで、基本的な音声合成の処理として録音編集方式とテキスト読み上げ方式の二つを紹介します。音声合成の仕組みを理解すれば、その違いをいっそう理解できると思います。

## ● 方式1：テキストを入力するだけで自然な文章をしゃべる「テキスト読み上げ方式(TTS)」

昨今では、文章を読み上げさせ、なるべく自然な発音にしようという技術がさらに進歩してきています。これはテキスト・トゥ・スピーチ(Text To Speech)と呼ばれ、略してTTSともいわれます。TTSはテキストを入力すると、ソフトウェアで音声を読み上げて出力します。パソコン上でアナウンスを作成したり、あるいはWebコンテンツなどのテキストを読み上げたりするのに使われています。極端な例としてはテレビのナレーションをすべてこのTTSによって生成している場合もあります。

このような技術は昨今のパソコンやスマートフォンなどで使われることが多くなっており、メールやWebページを

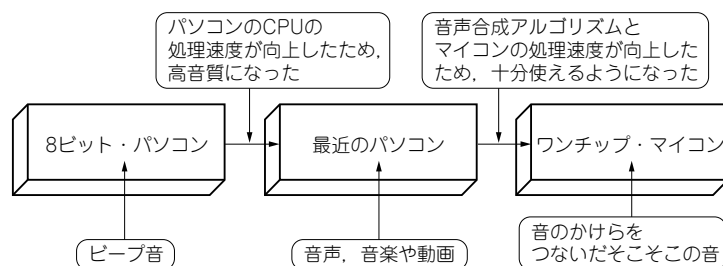


図1 ワンチップ・マイコンでもそれなりにちゃんとしゃべれる

マイコンの性能が上がり、コンパクトな音声合成処理アルゴリズムが普及してきたことで音声合成をマイコンでも扱えるようになった



写真1 テキストをシリアルで受信すると音声に変換するIC「ATU3011F4-PU」