

ラク楽実践 VR

- 手と足と頭を使え！ VR システムの作り方 -

第 30 回

音声合成チップ AquesTalk pico LSI

吉元俊輔（大阪大学）

1. はじめに

音声合成技術の発達によって、Google 翻訳における音声読上げ機能や VOCALOID による歌唱など、機械に“しゃべらせる”ことをインタラクティブにできるようになった。本稿で紹介する音声合成チップ AquesTalk pico LSI は、このような機能を一つの IC に搭載し、シリアル通信によって送られてきた文字をリアルタイムに読みあげてくれる、いわば“しゃべる”IC である。驚くほど手軽に利用できることから、VR 研究においても効果的なツールとして役立つことを期待し、その概要と使用体験について報告する。

2. AquesTalk pico LSI について

AquesTalk pico LSI を紹介する前に、その技術的背景を説明する。音声合成技術は古くから研究がなされており、実用化が進んでいる。1978 年に登場した Speak&Spell (Texas Instruments 社製) は、音声合成によって読み上げられる単語のスペルを学習する、子供向けの学習用玩具である。音声合成 LSI という当時では斬新な技術を導入しており、世間の注目を集めた。音声合成の最大の特徴は、録音データの再生のように大量のリソースを必要としない点にあり、LSI 上に機能を搭載できるということから様々な電化製品への導入が実現されている。近年では、ROM に登録された音声の再生だけでなく、ユーザの入力結果をリアルタイムに読み上げることが可能になりインタラクティブ性が向上したこと、アルゴリズムの改善によって発音の不自然さや音質が改善されたことによって、再び注目を集めている。

音声合成機能を LSI に搭載し、コンピュータから送信

された文字列の音声波形への変換を可能にしたのが音声合成チップ、AquesTalk pico LSI (アクエスト社製) である。AquesTalk pico LSI は図 1 に示すようなワンチップのマイコン (Atmel の ATmega328) に AquesTalk pico という日本語の規則音声合成ライブラリを縮小化して実装したものである。また、シリアル通信の機能を備えており、送られてきた文字列を音声波形に変換し、リアルタイムに再生する。



図 1 AquesTalk pico

一般的に音声合成でヒトの発声を実現するためには、文章に対して (1) 言語処理, (2) 韻律生成, (3) 波形処理を行なう必要がある。言語処理では読みや間・アクセントを、韻律生成ではリズムやイントネーションを決定する。さらに波形処理では音声波形データを滑らかにつなぎ合わせる。AquesTalk pico LSI においてもこの三つの機能が備わっているが、言語処理および韻律生成に関してはユーザが文章に記号を付加して設定しなければならない。これはプログラムサイズの縮小とリアルタイム処理を行うことを狙いとした実装であると考えられるが、文章を送る際には適切な位置に区切りやアクセントを入れて文字列を送信する必要がある。しかしながら、この作業を自動で行ってくれる音声記号列生成 Web サービスも提供されている。音声と文

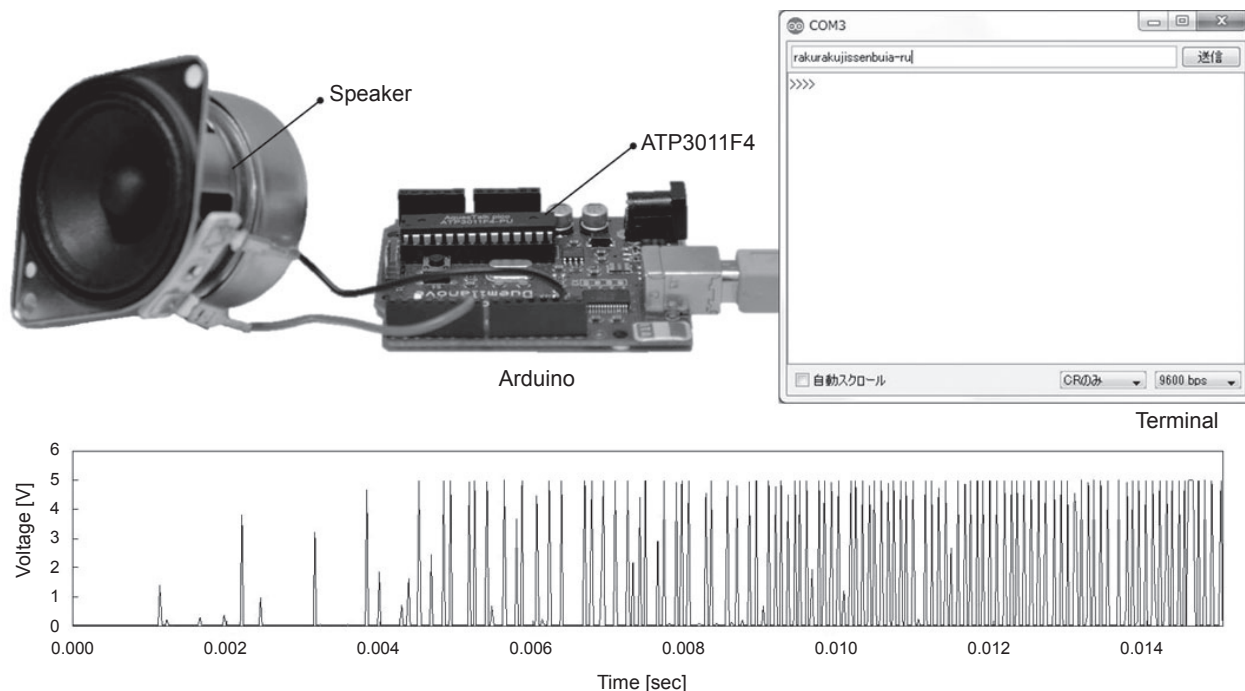


図2 音声合成 LSI による “ka” 音声再生結果 (サンプリング周波数 44.1kHz)

字・記号の対応や詳細な仕様についてはアクエスト社のホームページに詳細なデータシートが公開されているためそちらを参照されたい。

3. 音声合成 LSI の動作

Arduino Duemilanove を用いた最も簡単な使用方法について紹介する。音声の種類ごとに LSI が存在する (現段階では 5 種類) が、本稿ではそのうち「かわいい系の女声」と説明されている ATP3011F4-PU を使用する。まず、Arduino に組み込まれている ATmega328 を取り外し、AquesTalk pico LSI を装着する。AquesTalk pico LSI の出力ピンに対応する端子にスピーカを接続してハードウェアの準備は完了である。後は USB ケーブルでコンピュータに接続し、COM ポートへ文字列を送信するのみである。文字や文章をローマ字読表記で COM ポートに送信するだけで、文字通りの音声をはっきりと聞き取ることができた。図 2 にシステムの概要と、例として音声 “か” を再生した際の波形の一部を示している。アナログ出力ではなく PWM 出力によって音声波形を表現しているという点も、マイコンチップを利用した音声再生ならではの特徴である。

キーボードを叩いて送信した瞬間に再生される音声を聞き、思わず笑みがこぼれた。IC チップから音声が

インタラクティブに再生されるという状況は、なんとも言えない不思議な感覚を生じる。一部不自然さが残るものの、アクセントを指定して文章を読み上げさせると、確かに抑揚をつけることができた。手軽な利用が可能であるにもかかわらず、高品質な音声合成を提供可能であるという特徴を生かし、様々な機器への組み込みにより、擬人化を図るというのも面白い試みであろう。音声バリエーションの拡大や国際化など、今後の製品開発に期待が高まる。

4. おわりに

文字列の音声波形への変換を行う音声合成チップ AquesTalk pico LSI について紹介した。シリアル通信で送られてきた文字を軽快に発声してくれる、インタラクティブな音声合成が可能であることを確認した。非公式ではあるが、再生速度を変える隠しコマンドが存在し、音の高低を変えることも可能である。このような高品質な音声合成技術を手軽に利用できるようになった価値は大きい。Web を検索すると、AquesTalk pico LSI を用いて作成した興味深い応用例が動画を含め多数公開されている。学会員のみなさまの中から一味違った使い方が提案され、ユーザを驚嘆させるような音声合成システムが開発されることを期待する。