

音声認識 & 音声合成の キー・テクノロジー

山本 隆一郎

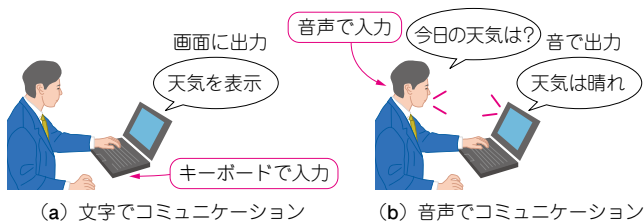


図1 コンピュータと人間のコミュニケーションはテキストで行われてきたが、音声で会話する技術が身近になっている

コンピュータの誕生初期から現在まで、コンピュータと人間は主に図1(a)のように文字(テキスト)でコミュニケーションしてきました。しかし、人は音声でコミュニケーションを取り、思考を積み重ねる動物です。そこで、図1(b)のようにコンピュータと人の間で、より自然にコンピュータと会話したいという欲求が生まれました。これを実現できる音声合成と音声認識という技術が注目されています。

本章では、音声認識と音声合成の原理を解説し、これを実現するためのライブラリやICを紹介します。



音声のデジタル化

音声を合成したり認識したりするのに最も重要なのは、音の記録と再生です。

19世紀のトーマス・エジソンの蓄音機の発明や、20世紀に発明された磁気記録式の録音再生装置で、人類は自由に音声を再生できるようになりました^{注1}。この音声再生の原理は、簡単な言葉を繰り返ししゃべるおもちゃの人形や、電話の自動音声案内、放送などで活用されています(記録媒体は異なる)。

テープ・レコーダをエンドレス・テープの構造にすれば、延々と同じ音声を再生できます。しかし、これはあくまで記録された音声を忠実に再現するだけなので、決まった言葉しか再生できません。そこで、より自由度の高い音声再生が求められました。

● 音声の記録・再生が簡単に！

これを可能にしたのが、アナログの世界の音声信号を

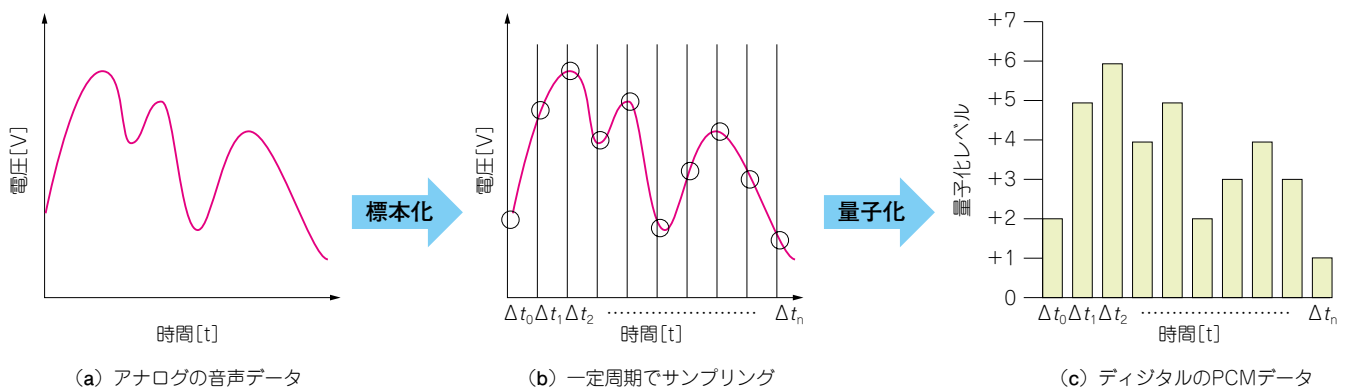


図2 音声をアナログ値からデジタルに変換し、PCMデータを得るとコンピュータで扱えるようになる

注1：人が人以外のものに音声を記録したり再生させたりしようとする試みは古くから行われていたようだ。18世紀に機械的に音声に近い音の再生に成功した記録がある。