

音声信号の基礎知識

ご購入はこちら

三上 直樹

最近、スマート・スピーカ (AIスピーカとも呼ばれている) が話題になっています。アマゾン、グーグル、LINEなどから1万円前後で販売されているようです。

スマート・スピーカに話しかけると、それにいろいろな応答や反応をしてくれます。例えば、スマート・スピーカと音楽ストリーミング・サービスを組み合わせれば、聴きたい曲をネットから探して流してくれたり、スマートIoT照明のPhilips Hue⁽¹⁾と組み合わせると、部屋の照明をコントロールしたり、「ルンバ」に指令を発したりと、いろいろなことが「声」を使ってできるようになってきました。

スマート・スピーカの「キモ」は、何とんでも音声認識の技術です。そこで、本稿では、音声認識で使われる音声情報処理と音声の生成や人間が音声聞き分けるメカニズムなどについて説明します。

なお、本文中で「音声認識」という用語を使いますが、これは「人間を介さずに、コンピュータのような機械で自動的に認識すること」という意味で使います。

人間はどのようにして音声を聞き分けているのか

私たちは音声を聞き分けて、何と言っているのか認

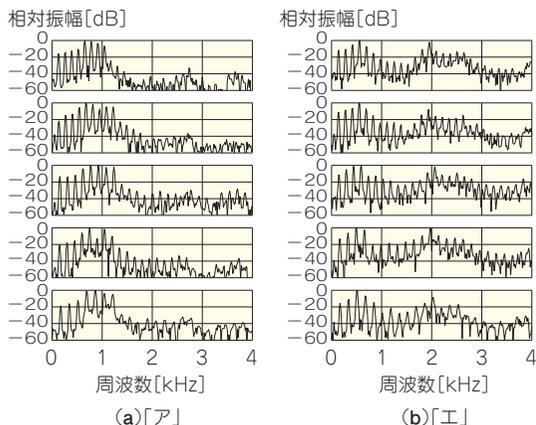


図2 母音の周波数スペクトラム波形は同じ音ならばほぼ同じになる
図1の波形をFFTして求めた

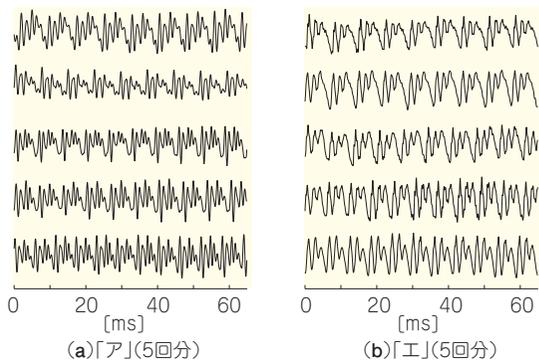


図1 母音の波形…多少波形が異なっても同じ音に聞こえる

識できます。音声認識はパターン認識の1分野です。音声に限らないのですが、人間はパターン認識という能力で、現在のところコンピュータよりも優れているといっただいでしょう。

音声認識の研究が始まったのはかなり古く、筆者の知る限りでは1950年代には既に始まっていました。しかし、それなりに実用的に使えるようになったのは、つい最近のことです。

音声認識を行うためには、まず私たちがどうやっ

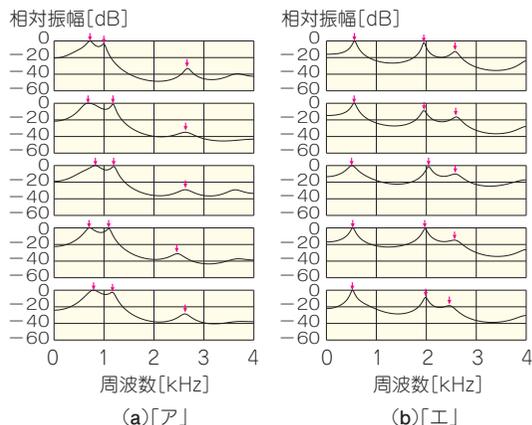


図3 同じ音ならばピーク周波数が同じになる…フォルマント周波数
図1から線形予測法で求めたスペクトラム。矢印はフォルマント周波数に対応するピーク