

Kinectのマイクロホン・アレーによる音声信号処理

鎌土 記良, 大沼 侑司, 猿渡 洋, 鹿野 清宏, 高橋 祐

Kinectは画像センシング・デバイスとして画像入力処理に注目されているが、マイクロホン・アレーを搭載しているので、音声センシング・デバイスとしても使用できる。ここでは、安価で容易に入手が可能なKinectを使って、音声信号処理を行う方法を紹介する。(編集部)

1. 音声センシング・デバイスとしての Kinect

Kinectは、画像センシング・デバイスとして非常にコスト・パフォーマンスの優れた画期的な製品です。実は、その内部に搭載されている音声センシング・デバイスも画像処理部分と同じようにコスト・パフォーマンスが優れており、画像と音声と同時に一定の品質で処理できる安価なデバイスとして、音声信号処理研究者からも大きな注目を集めています。

音声信号処理の研究者の視点から見た場合、Kinectが音声センシング・デバイスとして優れているところは2点あります。一つ目は汎用のUSBオーディオ・デバイスとして使用できること、二つ目は画像センサとのパッケージとなっているということです。

前述のように、KinectはUSBオーディオ・デバイスとして扱えるため、現在はWindowsに限られてはいるものの、汎用的なオーディオAPI [Microsoft Kinect SDK (以降MK-SDKと呼ぶ)のドライバではWASAPIやASIO] を使って音声センサの出力をそのままプログラム内ヘリアルタイムに取り込めます。そのため、Kinectに接続するパソコンの処理能力次第で、実環境で動作する音声信号処理アプリケーションを容易に作成することができます。

また、近年では、画像認識処理の結果を活かした高精度な音声信号処理手法の研究も行われており、画像と音声の複合センサ・デバイスを必要とするような研究においても、Kinectだけで対応できるという利点があります。

本稿では、このような優れた特徴を持つKinectについて、奈良先端科学技術大学院大学での応用事例を交え、Kinectによる独自の音声信号処理を効果的に実装する方法について述べたいと思います。なお、OSにはWindows Vista以降を、開発環境にはMicrosoft Visual Studio 2010 Express Edition以上を、Kinectの音声センサであるマイクロホン・アレー・ドライバにはMK-SDKに付属するものを使用することを前提として話を進めていきます。

2. オープン・ソース・ライブラリの活用

MK-SDKは、WASAPIを用いてマイクロホン・アレー出力をプログラム側で取得するサンプル・プログラムが付属しています。しかし、独自のマイクロホン・アレー信号処理をこのサンプル・プログラムに追加するためには、初学者には敷居が高いのが現実です。

そこで本稿では、Kinectを用いた独自の処理を実装するために、WASAPIに関する知識がなくても簡単にKinectをハンドリングすることのできるPort Audio (図1) というライブラリを用います。Port Audioは、OSやデバイスに依存しないさまざまなオーディオAPIのラップであり、高い移植性を備えています。また、複数台のKinectも簡単にハンドリングできるため、近年、盛んになってきている分散マイクロホン・アレー信号処理にKinectを応用できるスケーラビリティも持ち合わせています。さらに、Kinect SDKで提供されているサンプルではマイクロホン・アレー入力信号をwavファイルへ直接保存する保存コードが書かれていますが、こちらもlibsndfile⁽²⁾というライ