

KINECT™ テクノロジーと マイクロソフトが提案する NUI の可能性

千葉 慎二[†]

マイクロソフトが提案する新しいユーザーインターフェイスを実現する第一歩として誕生した KINECT。そこに至った経緯、KINECT の特徴や NUI との関係、ゲームをこえた分野への展開の可能性を紹介する。

KINECT™ Technology and Potentiality of NUI which Microsoft proposes

SHINJI CHIBA[†]

KINECT is the first step which realizes the Natural User Interface (NUI) which Microsoft proposes. This document is written the details of KINECT, a relation with NUI, and the potentiality of extent to the other field beyond the game.

1. ユーザーインターフェイスの進化と KINECT

コンピューターのユーザーインターフェイスは、キーボードを用いてコマンドプロンプトに入力するキャラクターユーザーインターフェイスから、マウスを用いて画面上のアイコンを選択するグラフィックユーザーインターフェイス、ペンを用いて文字や絵を直接入力するタブレット型ユーザーインターフェイスへと進化してきた。また、ここ最近ではスマートフォンやマイクロソフトサーフェスに代表されるタッチ型ユーザーインターフェイスが主流となっている。その次の進化形としてマイクロソフトはナチュラルユーザーインターフェイスを提案している。

ナチュラルユーザーインターフェイス（以下 NUI）には従来のインターフェイスと大きく異なる点がある。従来のインターフェイスは人間が機械の使い方を覚えてはじめて入力が可能となる方法であったのに対し、NUI は機械が人間の直感的な行動を読み取ってより自然な反応（出力）をさせることができるという点である。

電子機器のインターフェイスに対するユーザーの不安や悩みを調査したところ、操作方法がわからないので触れたくないといった意見が多かった。人間同士なら声や身振りでコミュニケーションが簡単に取れるのに、機械が相手だとそれができないといった内容である。

コンピューターを活用して人間の生活をより豊かにするための研究をマイクロソフトは長年行っている。ユーザーの声を理解し問題を解決するためにはナチュラルユーザーインターフェイスが最適であると考え、それを実現する第一歩として KINECT を開発した。

2. NUI を実現するための KINECT の機能

KINECT は買ってきてそこに置くだけで NUI を実現する魔法の箱というわけではない。中身は多数のセンサーで構成された電子機器である。その中に含まれる特徴的な機能とソフトウェア、そして優れたコンテンツをうまく組み合わせることによって NUI を表現することができる。

KINECT は Web カメラと同等の性能を持つカラーカメラと、赤外線レーザープロジェクターおよび赤外線を撮影する深度カメラ、撮影された赤外線の映像から深度を推定するプロセッサ、そして 4 つの指向性マイクアレイ、不要なノイズを除去するための DSP がビルトインされている。これらのカメラやマイクで収集したストリーム情報を USB 2.0 のバス経由で Xbox 360 あるいは Windows PC に転送する。

[†](株)日本マイクロソフト
Microsoft Japan Co., Ltd.

台座部分にはチルトモーターを装備しており、例えばカメラの設置箇所より低い位置にある物体や高い位置にある物体までの距離をカメラに触れることなく計測できる。また、KINECTには重力センサーも内蔵されており、傾いた場所に設置されてもチルトモーターによって常にカメラを正面に向かせることができる。

KINECTで注目される機能のひとつに骨格追跡や声の認識があるが、これらはハードウェアではなくソフトウェアで処理される。全部で20箇所の関節位置が特定でき、各関節は(X,Y,Z)のカメラを中心(0,0,0)としたミリ単位の実空間座標で提供される。ユーザーはそれらの情報を元に人の動きの情報を得るかたちとなる。

声の認識はあらかじめ入力される予定の言葉を事前に登録しておく方法をとる。また音声認識の際邪魔になるノイズをリアルタイムに除去する機能が備えられている。ここでいうノイズとは「人の声以外の音(例えばスピーカーから出力している音)」という意味以外にも「入力しようとする人以外の声」つまり「他人の声」もノイズと定義づけられる。音声認識では特定の人の声だけが受け入れられるべきなので、例えばビームステアリングによって骨格追跡している人の方向の音声のみにフォーカスし、それ以外の方向から発せられるノイズを抑制することもできる。

3. KINECTを用いたNUIの実現

NUIはその用途によって実装および表現方法が異なるため、こうすべきといった決まりはないが、KINECTを用いた直感的なNUIの実現方法のひとつにアバティアリングという手法がある。アバティアリングとは、KINECTの入力とアバター(自分の仮想的な分身)の動きを対応させることである。これを実現する方法はとても単純で、KINECTの骨格追跡情報を3Dアバターの関節にマッピングするだけである。ただ、その際にはいくつか気をつけなければならないことがある。

- ミラーリング
仮想空間のアバターは出力画面の上では通常、プレイヤー(KINECTのセンサーの前に立ってアバターを操作する人)に対して正面を向く。そうすると左右が逆転して映ることになる。その際、例えばプレイヤーが右手をあげたときアバターも同様に右手をあげるとプレイヤーは一瞬戸惑いを感じてしまう。この場合人間心理的により自然な入力と感じさせるようミラーリングした状態で出力したほうが良い。これはアバティアリングに限らず、例えばカラーカメラや深度カメラの情報をAR(拡張現実)表現で使用するケースでも同じことがいえる。
- スケーリング
人はそれぞれ背格好が異なる。したがって自分に似た体型にアバターを補正すべ

きである。特に複数人が仮想空間のアバターとして同時に表現される場合はなおさらである。ただしアバターが自分の分身ではなく仮想的な別人(例えば、ゲームの中のキャラクターなど)を演じる場合この点を考慮する必要はない。

- 自己交差テスト
骨格追跡情報には運動学が考慮されていない。そのためアバターを表現する場合、現実世界と整合性が取れないような動作を避けるよう逆運動学を用いて動きを補正あるいは抑制すべきである。そうしない場合、プレイヤーはアバティアリングの表現に違和感を覚える。

NUIの実現方法にはジェスチャーによる入力手段も有効であるが、その場合は特定のジェスチャーや特殊なジェスチャーを強要してしまわないよう注意する必要がある。人が入力のための動作を覚えるのではなく、人が直感的に動いた結果が想定通りの動きとなるようコンピューター側で理解することこそがナチュラルユーザーインターフェイスであるということを忘れてはならない。

しかし、漠然と入力を待つだけでは人は不特定な行動をとってしまう。それらの行動を包括的に区別することは非常に困難である。そのため人が想定された行動を自然にとるよう、人の行動を導くという手段もNUIを実装する上での重要なポイントになる。例えば「扉をあけて先に進む」という動作を、図1のようなドアノブがついた扉が目の前にあるケースと、図2のような引き戸が目の前にあるケースとで比べてみよう。図1を見た多くの人は次の入力を行う際にまず向かって右側についているドアノブに触れてドアを押そうとする(押して反応がなければ引こうとする)が、図2のような引き戸を見ると手を左から右に動かそうとする。どちらも「先に進む」ための動作であるが、入力インターフェイスをみた人の多くがとるであろう行動を、インターフェイスの作り方によってある程度導くことができるのである。

このように、その人がとる行動を入力インターフェイスの工夫によってその人さえも気づかないうちに誘導することでより自然さを感じさせることができるようになる。



図 1 ドア
Figure 1 Lever-style door



図 2 引き戸
Figure 2 Sliding door

利用目的によってはジェスチャー入力が必要なく、単にそこに人が存在しているかどうかということ自体も自然な入力となりうる。例えば、カメラ付近を歩く人物を捕捉したとき、その人物の近くにある壁掛けモニターのような機器に何らかの出力（表示等）をすると、たいてい人はそれに気がつき、それが何であるかを確認しようといったん足を止める。これは不特定多数の人に見てもらおうことを目的としたサイネージ用途に有効な手段である。その後、よりインタラクティブなジェスチャー入力へと導くことでサイネージの効果をさらにあげることができるかもしれない。

NUI を実現するにはフィードバックも重要である。特に KINECT 単体だけでは触覚フィードバックが（実際の物体に触れて反応があるようなケースを除いて）基本的に得られない。ほとんどの操作を空中でしなければならぬため、入力が正しい場合あるいは間違っただけの場合はプレイヤーに対して何らかの正当なフィードバックをする必要がある。そうしなければ人はそこに不安と不自然さを感じてしまう。

フィードバックには出力デバイスで何かを表現したり、音声を使う方法あるいは直接的なフィードバックデバイスを準備するなどが考えられる。例えば単純なところでは、パワーポイントをめくる動作で手を右から左へフリックするジェスチャー入力を作ったとすると、パワーポイントのページは同じく右から左へ移動してめくられるよう表現し、逆の動作を行った場合ページは反対に左から右にめくられるよう表現する。

NUI の実現には声による入力も欠かせない要素のひとつである。人の声は身振り手振りと同じくらい自然な入力手段であるが、音声認識においてもやはり操作方法を強要してはならない。そのとき人が自然に発するであろう言葉を理解し、それを即座に

反映させることが重要である。例えば「マイクロソフトのコンペティターは？」という問いかけに対しては「グーグル」や「アップル」などが反応すればとても自然だ。

さらにもっと自然な入力へと発展させるため、個々の使用者の行動や特徴を記録してその都度必要な情報を学習させていけば、より優れた入力へと変化していく可能性がある。そのようなアプリケーションおよびサービスが理想である。

4. KINECT と開発 SDK について

KINECT を Xbox 360 環境で使うための SDK は残念ながら一般に公開されていない。しかし Windows 環境で使うための開発用 SDK および開発環境は無料で入手可能である。32-bit と 64-bit 環境用の 2 種類のドライバーおよびライブラリーが存在するが、通常開発者はこれらの差を気にする必要はない。研究、ホビーユースまたは教育目的であれば無償ソフトやプロトタイプの実験やデモが可能である。ただし商用には別途ライセンスが必要である点に注意しなければならない。

5. KINECT の異なる分野への展開と可能性

KINECT は Xbox 360 ゲーム用の周辺機器として 2010 年 11 月にリリースされたが、現在ではゲームを超えた様々な分野での展開が注目されている。例えば、ロボットの目として、無菌状態が求められる医療現場での利用、リハビリテーション改善プログラムの支援機器、デジタルサイネージ、簡易的なモーションキャプチャー装置など様々な用途でソフトウェアの研究・開発がされている。アイデア次第で KINECT 利用の可能性はほぼ無限に広がると考えられ、マイクロソフトも KINECT の様々な分野における展開と NUI の発展に大きく期待している。