

ビジュアルプログラミング環境向け音声入力操作支援インタフェースの改良 - 音声訂正および音声コマンドの導入

田中 秀明[†] 平井 重行^{††}

Improvement of Speech Input Interface for Visual Programming - Implementation of Speech Repair and Speech Command

HIDEAKI TANAKA,[†] SHIGEYUKI HIRAI^{††}

1. はじめに

音声認識技術を活用した実用的なインタフェース研究^{[1][2][3]}の一つとして、我々は Max/MSP や PureData などのビジュアルプログラミング環境における音声認識を用いた操作支援について取り組んでいる^[4]。これらの環境では、ある機能を持つ箱（オブジェクト）線で接続し、データフローを視覚的に記述するスタイルでプログラミングを行う。操作は、マウスカードルでオブジェクトの配置や接続を行うが、オブジェクト名称などの機能割り当てにはキー入力が必要である。そのために片手をマウスとキーボード間で頻繁に移動させるオーバーヘッドが発生し、操作効率が悪くなることが多々ある。我々は、実際のプログラミング環境に対してオブジェクト名を音声認識する操作支援インタフェースを構築し、キー入力と音声入力による操作効率の違いについて実験を行ってきた^[4]。しかし、音声認識誤りによる別のオーバーヘッドが発生する問題があったため、今回は先行研究^[4]を参考に入力機能の操作性を向上するとともに、音声コマンドなど新たな機能を導入して入力インタフェースの改良を行った。本稿ではその実装について述べる。

2. 音声入力インタフェースの改良

2.1 入力インタフェースのコンセプト

我々の音声入力インタフェースでは、既存の音声認識エンジン(Julius)を利用し、認識対象をオブジェクト名や数値に絞った辞書を用意することで、目的の入力支援環境を構築している。しかし、音声認識誤りを考慮していないため、謝った文字列を入力してしまうこともあった。この音声認識誤りについては、緒方らの研究^[4]で、競合候補をリスト提示してユーザに選択させる手法が提案されており、我々は同様の手法を取り入れることとした。

一方で、言い間違いなど候補リストに選択すべき

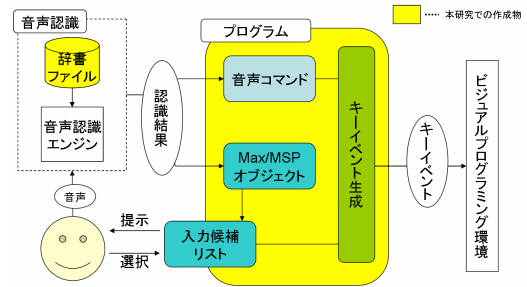


図 1. システム概要

文字列が含まれていない場合には、リスト表示をキャンセルする必要がある。また、プログラミング上必要なショートカットキー操作でオーバーヘッドとなるものもある。そこで、これら文字列入力以外の操作は別途「音声コマンド」として音声認識対象とし、特定の処理を行う機能も追加した。なお、音声コマンドはオブジェクト名とは重複しないものを設定する。これらにより、音声認識誤りやプログラミング環境の操作性を考慮した入力インタフェースを実現する。

2.2 処理の概要

今回は Windows 用 Max/MSP に対する音声入力インタフェースとして実装している。その処理構成を図 1 に示す。入力された音声から音声認識エンジンを用いて音声認識を行い、認識結果からオブジェクト名(文字列)か音声コマンドかを判断する。文字列入力であった場合、マウスカードル位置に認識結果の上位候補リストを提示し、ユーザが選択した文字列をキーイベントとして生成し、Max/MSP へ送る。また、音声認識結果が音声コマンドであった場合、コマンドに応じて候補リストをクリアしたり、対応キーイベントを生成して Max/MSP へ送る処理を行う。各機能の詳細は次章に示す。

3. 各機能の処理と実装

3.1 音声認識機能と辞書データの作成

本研究では音声認識エンジンに Julius を用いており、オブジェクト名や音声コマンドの認識を行うために Julius 用辞書ファイルの作成を行った。辞書データ自体は「OBJECT [metro] m e t o r o」のような形式で、主に単語認識が主な内容であるため N-gram 形

[†] 京都産業大学理学研究科

[†] Graduate School of Division of Science, Kyoto Sangyo University

^{††} 京都産業大学理学部

^{††} Faculty of Science, Kyoto Sangyo University

式の辞書にはしていない。現時点ではオブジェクト名が約 150 種、記号 36 種、アルファベット 26 種、0～999 までの数字、および音声コマンド数種分の辞書データを用意している。

なお、GUI オブジェクトに関しては、通常はプログラミング時はウィンドウ上部に表示されるオブジェクト・パレットから目的オブジェクトを選択するのだが、実は普通のオブジェクトボックスに GUI オブジェクトの名称をキー入力することでもオブジェクト生成が可能である。そこで、それら名称も辞書に登録することで、音声で GUI オブジェクトの生成も可能となっている。これは従来のマウス選択による生成よりも操作ステップ数を減らすという利点がある。

3.2 競合候補リスト表示処理

認識結果がオブジェクト名や数値などの文字列入力を意味するものである場合、最大 5 個の競合候補をマウスカーソル位置にリスト表示する(図 2 参照)。これは先行研究^[4]が提案する「音声訂正」機能を実装したもので、音声認識誤りは常に起こりうることを許容するインタフェースとして有効である。ここでは、表示されたリストからユーザが選択した候補文字列に対し、各文字をキーイベントに変換して Max/MSP へ送信する。なお、Julius で複数の認識候補を得るためには N-best パラメータの設定をしておく必要がある。

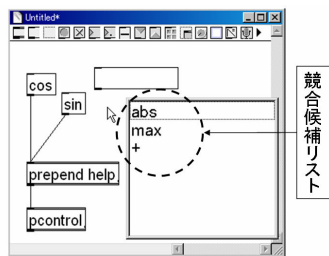


図 2. 競合候補リスト

3.3 音声コマンド処理

現時点で処理可能な音声コマンドの読みおよび動作を表 1 にまとめる。「キャンセル」は状況に応じて 2 種の動作がある。1 つは候補リスト表示中の状況で、音声認識誤りや言い間違いなどが理由で選択すべき候補がない場合にリストのクリアを行う。もう 1 つはリスト選択を誤ってしまった状況で、直前の入力文字列分バックスペースを Max/MSP へ送信する。「オブジェクト」「メッセージ」「コメント」は作業ウィンドウの左上に存在している各ボックスを表しており(図 3 参照)、コマンド入力時には、マウスカーソルを各ボックスの座標に強制移動、クリック、元の座標に強制移動、を行うことで実現している。通常ならマウス操作でその作業を行う必要がありが、音声入力により、即座にカーソル位置へ

表 1. 音声コマンド例

種別	機能 (発話例)	動作
音声入力機能に関するもの	キャンセル	1.表示中のリストを消す 2.直前に入力した物を消す
	スピーチオン/オフ	音声入力機能のオン/オフ
Max/MSP の機能に関するもの	オブジェクト	オブジェクトボックスの生成
	メッセージ	メッセージボックスの生成
	コメント	コメントボックスの生成
	ショウ/ハイド	選択オブジェクトを表示/非表示対象に設定

のオブジェクト生成が可能となっている。ショートカットキーは単純に各キーイベントを生成して送信する。「ショウ」「ハイド」は、Max/MSP では頻繁に利用される機能だが、操作キーが右手で操作する位置にあるためマウスとの手のやりとりでオーバーヘッドとなる操作の代表例であることから操作効率を上げるため音声コマンドの対象とした。

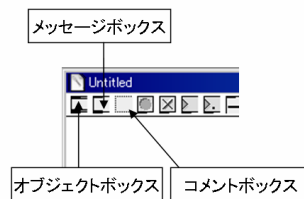


図 3. オブジェクト・パレット

4. まとめ

本研究では、Max/MSP 系のビジュアルプログラミング環境において効率よくプログラミングが出来るよう、音声によるオブジェクト名などの文字列入力や、様々な音声コマンドの機能を持つ操作支援インタフェースの開発を行った。今後の課題は、辞書データのさらなる充実や、本インタフェースによる操作効率の定量的・定性的評価を行うことが挙げられる。

参考文献

- [1] Steve Whittaker, Brian Amento, “Semantic Speech Editing” Proc. CHI2004, pp.527-534, 2004.
- [2] J. Ogata and M. Goto, “Speech Repair: Quick Error Correction Just by Using Selection Operation for Speech Input Interfaces” Proc. Interspeech 2005, pp.133-136. 2005.
- [3] Kazutaka Kurihara, Masataka Goto, Jun Ogata and Takeo Igarashi, “Speech Pen: Predictive Handwriting based on Ambient Multimodal Recognition” Proc. CHI2006, pp.851-860, 2006.
- [4] 田中秀明, 平井重行: “音声入力を用いたビジュアルプログラミング環境の操作率向上”, インタラクシオン2007 論文集, 2007.