

音声入力による ビジュアルプログラミング環境操作支援インタフェース

平井重行[†] 田中秀明[‡]

[†] 京都産業大学 コンピュータ理工学部

[‡] 京都産業大学大学院 理学研究科

ビジュアルプログラミング環境は一般的にマウス操作が重要な操作内容であるが、Max/MSP や PureData 等のタイプの環境では、オブジェクト名の割り当てなどキー入力が必要とする操作も頻繁に起こる。プログラミングに慣れてくるとマウス・キーボード間の手の移動がプログラミング作業のオーバーヘッドとなることから、我々はそのような環境での操作効率向上を目指して音声入力をを用いた操作支援インタフェースの開発を行っている。ここでは、オブジェクト名の文字列を音声入力する機能を基本とし、音声認識誤りを許容するための競合候補リスト提示機能、特殊操作を実行する音声コマンド機能などを実装している。本稿では、Max/MSP 向け操作支援インタフェースの各機能と実装について述べ、操作効率に関する被験者実験の結果を報告する。

Speech Input Interface for Visual Programming Environment

Shigeyuki Hirai,[†] Hideaki Tanaka[‡]

[†] Faculty of Computer Science and Engineering, Kyoto Sangyo University

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

In general, Visual Programming Environments require mouse operations mainly. Some environments, for instance Max/MSP, PureData, require also keyboard operations to input object names, parameters. In these environments, there are overheads by moving hands between a mouse and a keyboard for experienced programmers. We have developed a speech input interface for efficient programming. This interface has basic function to input object names by speech, and some functions for instance to display results of speech-recognition, various speech command for some operations. This paper describes these functions and an implementation of the interface software, and also results of experiments for efficiency.

1. はじめに

ビジュアルプログラミング環境の多くは、様々な機能を持つビジュアルなオブジェクトを GUI 画面上で配置し、その位置関係や線による接続などで処理の流れや関係性、論理などを表現してプログラミングを行う。ビジュアルな表現方法には様々なタイプのものが存在し、それらのうち Max/MSP や

PureData (Pd), VVVV 等はオブジェクトボックスに対してテキストで機能割当てを行い、ボックス同士を線で接続するタイプのものと言える。これらの環境ではオブジェクトの機能割当てのためにキーボードによる文字列入力が必須であり、マウスなどポインティングデバイスによる操作が中心となるビジュアルプログラミング環境の中では他のタイプのものよりもキー入力の頻度が高い。そのため、マウス

操作を行う手がキーボードとの間で何度も往復し、操作に慣れた人ほどその動作がプログラミング作業上のオーバーヘッドとなる。

我々は、この操作上のオーバーヘッドを減らし、より効率よくプログラミングするために、音声入力を用いて操作支援を行うインタフェースを提案し、音声入力の有効性を実験するとともに Pd と Max/MSP 向けに実装してきた¹⁴⁾。本稿では、Max/MSP 用に開発した操作支援インタフェースの機能について説明し、幾つかの機能について評価を行ったのでその結果について述べる。

2. 操作支援インタフェースの方針

操作支援インタフェースの基本機能は、キーボードの代わりに音声認識による文字列入力を行い、オブジェクト機能割当てや、パラメータの数字などを入力するものとした。この基本機能に加え、ショートカットキー操作やマウス操作、それらの組合せで行うプログラミング上の操作についても、操作支援が有効もしくは可能なものについては、音声コマンドとして入力できるようにする。ただし、キー入力やマウス操作のすべてを音声入力に置き換えようとするのではなく、状況に応じて既存の操作と併用してマルチモーダルな入力インタフェースを実現し、プログラミングの操作効率を向上させることが目的と考えている。なお、オブジェクト間のパッチコード接続操作については、現時点では操作支援対象に含めていない。そのための操作支援ツールは既に存在しており¹⁵⁾、本研究のソフトウェアと併用すればよいと我々は考えている。

一方で音声認識の機能については、認識誤りが起こることを想定し、緒方らが提案する音声訂正¹⁴⁾を参考に、競合候補がある場合には複数候補を提示してユーザに正しいものを選択させる。ユーザが意図する入力を選択できることで、認識誤り時のフォローを迅速に行い、操作効率を落とさないようにすることが狙いである。

また、音声認識エンジンを含む操作支援インタフェースのソフトウェアは、フロントエンドとして実行させ、Max/MSP などのビジュアルプログラミング

環境には一切手を加えないという方針で開発を行ってきている。

3. 各機能の内容と実装

本研究の操作支援インタフェースが持つ機能とその内部処理は図 1 のようになっている。

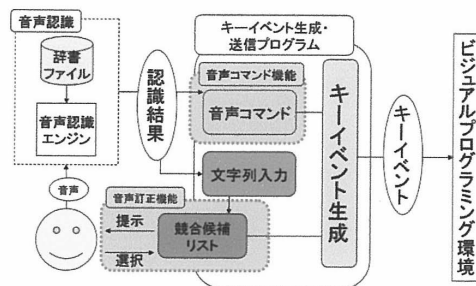


図1 操作支援インタフェースの動作概要

ここでは、マイク入力された音声を生認識エンジンでオブジェクト名などの認識を行い、その認識結果が音声コマンドであれば、そのコマンドに応じた適切な処理を行うか、ビジュアルプログラミング環境に対してショートカットキーなどのキーイベントを送信する。一方、音声認識の結果が、音声コマンドではない場合には、オブジェクト名や数字などとして扱うことになる。この場合は音声認識誤りが起こる可能性があることから、認識結果の候補リストをマウスカーソル位置にダイアログで表示し、ユーザに選択させる。選択された結果については、文字列に相当するキーボードイベント列を生成して、ビジュアルプログラミング環境へ送信する。

以下に、Max/MSP 向けに開発した操作支援インタフェースの各機能の詳細と実装について述べる。

3.1 基本的な音声認識機能

音声認識エンジンは既存のものを利用し、認識すべきオブジェクト名や数字、音声コマンドなどを記述した専用辞書ファイルを作成することで、本研究向けの専用認識エンジンとして動作させる。現時点では認識エンジンとして Windows 環境向けの Julius for SAPI 2.3 を利用している。

なお、次節に示す機能のために、Julius には N-best の設定を行って、複数の認識候補を出力するようにしている。

3.2 競合候補提示選択機能

似たようなオブジェクト名や発音の仕方によって音声認識誤りが起こりうることから、音声認識結果は常に上位複数候補を出力するようにしている(N-best)。この候補リストをマウスカーソル位置にダイアログで提示し、ユーザに意図した入力内容を選択させる。即時訂正を可能とすることで、認識誤りを許容すると共になるべく操作効率を下げないようにしている。競合候補リスト提示例を図2に示す。

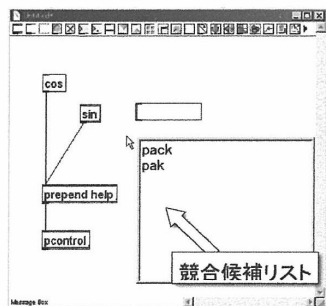


図2 競合候補リストの提示
(音声で「ぱっく」と発話した場合)

また、複数候補を提示する際には、趙らの研究^[5]を参考に、認識スコア値により提示候補数を絞り込むと共に最大5個までを提示する条件とした。なお、現時点では入力単語毎に認識結果がリスト提示されるようになっているが、オブジェクトの種類毎に取り得るパラメータの種類を、Juliusの文法ファイルとして適切に設定することで、本来の音声訂正^[4]に近い機能を実現する予定である。

3.3 音声コマンド機能

音声入力機能のオンオフ切替や、競合候補リストのキャンセルなど、インタフェースの動作自体に関する機能のほか、Max/MSPで利用できるショートカットキーの音声入力機能、Max/MSPではショートカットキーが割り当てられていない操作などについて、音声コマンド機能として動作するよう操作支援インタフェースを実装した。実装済みコマンドを表1(本稿末尾)に示す。なお、これらはオブジェクトとしては存在しないものなので、オブジェクト名入力の妨げにはならない。

以下に音声コマンドの種別毎に機能の説明を行う。

3.3.1 音声入力機能に関するコマンド

コマンドとしては、「キャンセル」と「スピーチオン」「スピーチオフ」の3種類がある。

「キャンセル」コマンドについては動作形態が2つあり、そのうち1つは提示中の競合候補リストにユーザが意図するものがない場合にリスト提示自体をキャンセル(画面上から消す)するコマンドである。もう一つは、直前に音声入力してビジュアルプログラミング環境へ送信されてしまったオブジェクト名などの文字列を削除するコマンドである。後者については、直前に送信した文字列の文字数ぶんバックスペースを送信する。

「スピーチオン」「スピーチオフ」は、音声入力機能自体のオンオフを切り替えるコマンドで、プログラミング中に本インタフェースのオンオフが随時できるようにしている。

3.3.2 ショートカットキーに関するコマンド

ショートカットキーは、Max/MSPで標準的に利用できるものはすべて音声コマンドとして実装した(表1参照)。マウスを非持ち手(多くの場合は左手を想定)で操作できるものは音声コマンドのほうが時間のかかることが推察されるが、マウスの持ち手が必要となるコマンドには有効となる場合もあると考えた。例えば、マウスの右手が持ち手の場合、「ショウ」「ハイド」は各々Ctrl-l, Ctrl-kが割り当てられており、常識的にはマウスからの手の移動が起こると想定される。

3.3.3 その他Max/MSPの操作に関するコマンド

Max/MSPで標準的にショートカットキーが割り当てられていない操作についてコマンドを割り当てた。現時点では、画面上に数多く配置するオブジェクトボックスやメッセージボックス、コメントボックスなど、プログラミング上必要なパーツの生成に関するものを挙げている。それらを生成には、図3にあるような画面左上のオブジェクトメニューからマウス選択するか、既存オブジェクトボックスなどをコピーして機能の再割り当てをするしかない。各々の生成コマンドを「オブジェクト」「メッセージ」「コメント」として割り当てた。

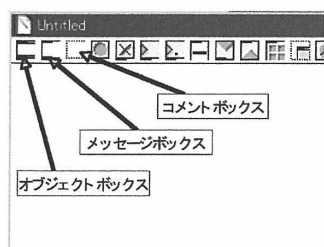


図3 オブジェクトメニューの各ボックス選択位置

これらの内部動作は、コマンドが認識された時点のマウスイカーソル位置から、各ボックスのメニュー位置へ内部処理としてマウス位置を移動させ、選択し、元のカーソル位置へ移動することでボックス生成を行っている。ユーザにはその動きは見えず、カーソル位置にボックスが生成されたように見える。

4. 操作支援インタフェースの評価

実装した操作インタフェースの競合候補提示機能と幾つかの音声コマンド機能に対し、それらの有効性や有効に機能する条件を確認するため被験者による評価実験を行った。本章ではその内容と結果について述べる。

4.1. 競合候補提示機能の評価

音声認識誤りが起こることを想定し、競合候補提示機能の有効性を調べるため、リスト提示の有無によるオブジェクト機能割当て速度の違いを計測した。

実験タスクとしては、被験者に画面上に生成・配置するオブジェクトを被験者に提示してから、1分間で可能な限りオブジェクトを生成（配置場所は自由）してもらう。この1分間の作業をリスト提示の有無で各5回実施した。なお、被験者は10名（男性9名・女性1名）である。実験前に目的と内容の説明を行ったのち、各被験者には10分程度の練習時間を与え、Max/MSPの操作や音声による文字列入力、提示されたリストからの選択など、一通りの操作は可能な状態にしてもらった。これにより作業の慣れを排除したうえで実験を行った。図4に1分間あたりのオブジェクト生成個数の平均と標準偏差を示す。

この図より、平均値の比較から競合候補リスト提示のあるほうがより多くのオブジェクトを生成して

いることがわかる。また標準偏差についてはリスト提示あるほうが少なく、生成時間にばらつきがないことがわかる。これらのことから、競合候補提示機能のあるほうが、より早くオブジェクトへの機能割当てが行え、作業時間の安定することが確認できた。

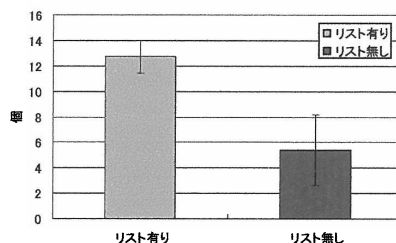


図4 競合候補提示機能の有無によるオブジェクト生成個数

4.2. 音声コマンド機能の評価

音声コマンド機能の評価を行うために、既に入力した文字列の取り消しを行うキャンセルと、オブジェクトボックス生成の音声コマンドについて評価実験を行った。ショートカットキーに関しては、音声入力に関する基礎実験^[1]における1文字入力と同じ実験であることから本報告では省略する。

なお、以下の評価実験では、いずれも被験者は12名（男性11名・女性1名）であり、各実験前の練習などの条件は前節の実験と同様である。

4.2.1 キャンセルコマンドの評価

オブジェクトボックスに入力されたオブジェクト名などを削除する場面を想定して、音声コマンド「キャンセル」による削除と、キーボードで文字数ぶんバックスペース入力による削除の2つの方法で作業時間の比較を行った。具体的には、キーボードでの削除作業は3, 5, 8, 11, 14文字のオブジェクトについて各10回行い、音声コマンドでの削除についてはオブジェクト名の文字数に依存しないので誤認識の起きにくいものをあらかじめ指定して10回行った。この実験結果を図5に示す。このグラフでは音声コマンドによる時間を横線で表現している。

この結果からは、キーボードでの作業時間は文字数に応じて長くなることがわかる。また標準偏差の

値も文字数に応じて大きくなっているが、これは被験者間の差異によるものが大きい。

音声コマンドでの作業はキーボードでのキャンセル作業のおよそ 11～12 文字分にあたることから、音声コマンドでの文字列削除は、オブジェクトボックス内に 12 文字以上の文字列がある場合になると判断できる。

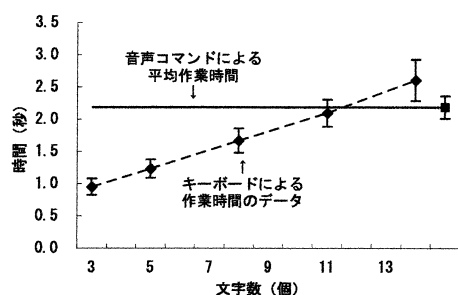


図5 キーボード入力と音声コマンドによる文字列削除にかかる時間の比較

4.2.2 オブジェクトボックス生成機能の評価

音声コマンド「オブジェクト」によるオブジェクトボックス生成・配置作業と、マウス操作によるオブジェクトメニュー選択・配置作業とで、作業時間の比較を行った。マウスでの作業は、オブジェクトメニューからのカーソルの距離に依存してマウス移動時間が変化することが想定できる。そこでマウスでの配置作業は、ウィンドウ上の初期座標値を A(160,120)、B(320,240)、C(480,360)、D(640,480) の 4 カ所 (図 6 参照) とし、各々の位置からオブジェクトメニューに移動し、元の位置にボックス配置を行うものとした。音声コマンドではマウス位置は実質的に意味が無いため、ウィンドウの任意の位置に生成・配置するものとした。被験者には、マウス作業の 4 つの位置および音声コマンドでの作業を各 10 回ずつ行ってもらった。実験結果を図 8 に示す。

図 7 の結果からは、マウスカーソルの位置がオブジェクトメニューからの直線距離でおよそ 800 pixel 以上ある場合に、音声コマンドでの作業の方が早いと言える。また、標準偏差はマウス移動のほうがどれほど大きい、これは被験者間の違いによら

ない。よって音声コマンドのほうが安定して作業できることが確認できた。

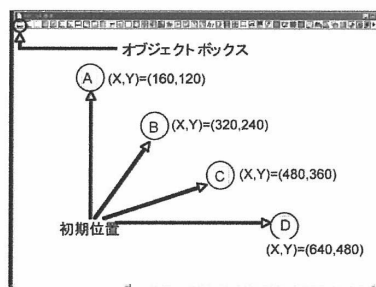


図6 オブジェクトボックス生成実験に関するマウス初期位置

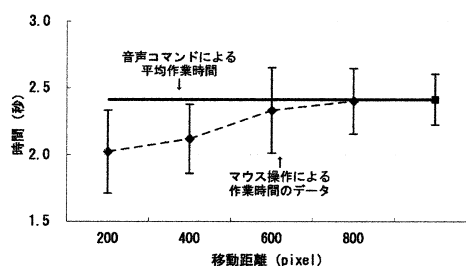


図7 オブジェクトボックス生成に関するマウスポインタ移動距離と時間の関係

4.3. 評価まとめ

競合候補提示機能の評価では、機能がなかった場合に比べてあるほうが倍以上の速度でかつ安定してオブジェクト名入力が可能である結果となっており、この機能が非常に有効に働くことが示された。

次に、音声コマンド機能の評価では、「キャンセル」による文字列削除について、オブジェクト名などの文字数が 12 以上の場合に音声コマンドが有効であることが確認できた。しかし、これは音声認識のレスポンスにも依存する結果であり、今後、より高速な PC 上で作業を行えばこの条件は改善が見込まれる。また、オブジェクト生成に関する音声コマンドの実験では、マウス位置がオブジェクトメニューからおよそ 800pixel 以上離れている場合に有効であることが確認できたが、これも同様に今後のより高性能な PC 環境で実行することで改善が見込まれる。一方で、オブジェクト生成は Max Toolkit⁴⁾

でショートカットキーが設定されているため、そちらを使うほうがよいことも考えられる。

5. おわりに

本稿では、音声認識技術を用いたビジュアルプログラミング環境向けの操作支援インタフェースについて、実装した機能の紹介と共に、作業効率に関する評価実験の結果について述べた。評価により機能毎の有効性や、有効に作用する条件が確認できた。今後は、ユーザインタフェースを改良してより使い易くすると共に、性能向上を目指してよりよい操作支援ができるインタフェースを目指す。

参考文献

- [1] 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効率向上, インタラクション 2007 論文集 CD-ROM (2007)
- [2] 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支援, インタラクション 2008 論文集 CD-ROM (2008)
- [3] Max Toolbox:
<http://code.google.com/p/maxtoolbox/>
- [4] 緒方淳, 後藤真孝: 音声訂正: 選択操作による効率的な誤り訂正が可能な音声入力インタフェース, 情報処理学会論文誌, Vol. 48, No. 1, pp. 375-385 (2007)
- [5] 趙國, 宮山章子, 山下洋一: N-best 音声認識における認識スコアを利用した候補提示数の決定, 電子情報通信学会論文誌 Vol. J88-D-II, No. 6, pp. 1003-1011 (2005)

表 1 音声コマンド一覧

種別	コマンド (発話内容)	動作
音声入力機能に関するコマンド	キャンセル	1. 表示中の競合候補リストのキャンセル 2. 直前に入力した文字列を消す
	スピーチオン/スピーチオフ	音声入力機能のオン/オフ
ショートカットキーに関するコマンド	パッチャー	Ctrl+N
	オープン	Ctrl+O
	クローズ	Ctrl+W
	セーブ	Ctrl+S
	プリント	Ctrl+P
	アンドゥ	Ctrl+Z
	リドゥ	Ctrl+Y
	カット	Ctrl+X
	コピー	Ctrl+C
	ペースト	Ctrl+P
	デュプリケート	Ctrl+D
	セレクトオール	Ctrl+A
	ファインド	Ctrl+F
	ファインドネクスト	Ctrl+G
	リプレース	Ctrl+=
	リプレースアンドファインド	Ctrl+J
	リジューム	Ctrl+R
	エディット	Ctrl+E
その他 Max/MSP の操作に関するコマンド	ショウ	Ctrl+L
	ハイド	Ctrl+K
	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

[6]