

音声による色表示平面内のカーソル位置制御

成田 健睦, 笈 一彦
Takechika Narita, Kazuhiko Kakehi

中京大学
Chukyou University
h20902m@st.chukyo-u.ac.jp

It is necessary to investigate the characteristics of speech utterances to design a system controlled by the sequence of speech utterances. First, the experiment of controlling a cursor by speech on a LCD, where a two-dimensional colour space (brightness was kept constant) was exhibited, was conducted. In the second experiment, subjects listened to the sequence of utterances in the first experiment, and then controlled the cursor due to it. The two experimental results suggest that the requirement for those systems controlled by speech : ① the system should use the prosodic information to distinguish a control utterance to the system from a utterance addressed to the speaker himself, and ② the system should make a confirmation inquiry to the speaker to desolve the ambiguity of meaning, especially for the word attributed to the characteristics of the space.

Keywords — speech, control, constraints free, confirmation, prosodic information.

1. はじめに

現在、カーナビゲーションシステムやアンドロイド携帯を初めとした音声認識機能が搭載された機器が数多く発売され、有効に使用されているケースも多い。それらの多くは検索目的に使用されるもので単語に対する音声認識がある程度正確であれば、実用上の問題が達せられる場合が多い。しかし、それらの機能が提供されていても、有効に利用されていない場合が多い。実際に使用する場面を考えた場合、運転中のカーナビ操作や携帯電話の発着信等手足を他の仕事に使用している状

態で認識させたい場合など機械に対しての操作のための特別な発話をしているという意識なしに自然な発話で操作できることが望ましい。制御や操作を目的とした場合には単なる音声認識だけではなく発話者の意図を正確に読み取ることが必要となる。

本研究では操作者が機械に対して音声による制御を行った場合、どのような発話特性が生ずるのかについて検討を行い実際に操作者が使えると感じるようなものにするために機械が備えるべき条件などを検討した。

2. 実験

2-1 実験概要

パーソナルコンピュータ(以下PC)の画面内において2次元のグラデーションのかかった色画面内から別途表示された指示色(ターゲット)と同等の色となる場所にカーソルを音声で操作する。実験時に使用した画面を図1に示す。



図1 実験時使用画像

図1の右端に表示されている部分が指示色である。画面右下にある「新規」と書かれている部分の色

を指示色と同じになるようにカーソル操作をする。カーソルは画面中央に表示されている(+)。カーソルの中央部分の色は右下隅に表示されカーソルが移動することにより新たにカーソル位置の部分の色が新規の色として表示される。

被験者が音声によるカーソル操作を行う実験 1 と実験 2 で行った各被験者の音声発話に従って画面上のカーソルを被験者に操作させる実験 2 を行った。

実験 1 では実際の音声認識ソフトは用いず、別室にて実験協力者が被験者の発話を聞き取りカーソル位置の操作する Wizard of Oz 法をとった。被験者には実験協力者が操作していることを伝えず機械が行っていると伝えた。また、実験協力者は被験者に与えられた指示色(ターゲット色)を知らず、音声入力のみによってできるだけ機械的にカーソルを操作するように指示した。

実験 1 と実験 2 では次のような相違がある、実験 1 では発話者自身の発話によって「機械」が動かしたカーソルの移動量や方向等の反応から次の発話内容を機械の操作特性に適応して決定していくことが出来る。また、裏で操作を行っている実験協力者も機械的な振舞いをするように指示されていても、被験者のこのような行動の構図は理解しているといえる。一方、実験 2 においては被験者がカーソル制御を行っても次の発話はその動きとは無関係である、即ち発話者側の適応的行動が無くなった場合に実験 1 と実験 2 に操作がどのようになるかを見るのが一つの目的である。

2-2 実験手順

本実験では両実験とも図 1 の左側部分は Microsoft excel 内の塗りつぶしの色選択画面を利用している。また同図右側部分は word を別途起動しセル内に塗りつぶしの色を表示することで指示色とした。図 1 の画面は色合い(色相:横方向)鮮やかさ(彩度:縦方向)がそれぞれ 0~255 までで数値化されている。表示画面における明るさ(明度)はどの色の部分でも同じとなっている。

実験 1(音声操作):

[被験者] 大学生及び大学院生 18 名

[実験課題] カーソル内の色と指示色が同等になったと感じるまでグラデーション画面内のカーソルの操作を音声によって行ってもらう。被験者には日本語による自然な発話により音声操作可能であると伝えた。実験では、以下の 3 群を設けた。

I. 「色属性制御群」: 色用語によるカーソル制御が可能
 II. 「位置情報制御群」: 位置・方向用語によるカーソル制御が可能
 III. 「自由発話群」: 使用する用語に対する制約を設けず自由な発話によるカーソル制御が可能

それぞれの実験群の被験者に対し指定された用語以外の発話では動作しないと伝えたが、その他の発話制限は行わなかった。即ち指定された用語以外の発話であってもカーソルは動かないので、いわゆる「不要語」を発話しないように特に気を使った発話をする必要が無いようにした。

実験協力者が被験者の発話をリアルタイムで聞きとるために、実験協力者と被験者の双方で Google トークとヘッドセットを使用した。実験協力者は実験中の被験者の手の動きや視線などヘッドセットから聞こえる音声以外の情報を得ることはできず、音声発話のみによる操作を行っている。実験 1 ではすべての状況をビデオにより録画し、音声も同時に収録した。実験協力者は被験者からは見えない別室にてマウスを使用してカーソルの操作を行った。

なお被験者・実験協力者共に同サイズのモニタを使用した。実験協力者が使用する画面では、図 1 の画面左側に示されている指示(ターゲット)色のカラーバーを隠して、被験者に提示している指示色が実験協力者に分からないようにした。また被験者が使用する画面では、通常画面下部に表示されるカーソル位置の色空間の数値情報を黒い紙片で隠した。

実験 2(実験 1 の音声による操作)

[被験者] 大学生及び大学院生 10 名

[実験課題] 実験 2 の被験者は、実験 1 において

録音した各被験者の操作音声を再生しその音声に従ってマウスによるカーソル操作を行った。被験者に対してカーソル動作に対する制約は設けず思った通りに操作するよう求めた。また、実験で使用する音声については事前にチェックを行ったが、制御用語とは関係ない独白があってもそれらは削除せず実験を行った。

実験には実験1で実験協力者が使用したPCを使用した。また、画面サイズによる感覚の違いを避けるため実験協力者が実験1のときに使用したディスプレイを用い実験を行った。実験2の状況はビデオにより録画をした。

3. 結果と考察

実験1における発話の分析から以下のような点が明らかとなった

操作者の発話の特徴：

ほとんどの操作者は、機械が対応していると考えていたが(丁寧な)依頼表現を使うもの(3人)と単に操作用語のみをいうもの(15人)とに分かれた。

操作者は自分の行為確認のための発話を多く行っている。①目標への到達の程度、②制御用語に対する「機械」の反応の確認、③制御プランの確認などである。

undoの発話は少ない(全発話233中2例)これが音声指示の特徴であるかについては、検討が必要である。

属性制御、位置制御、属性・位置制御の特徴：属性で制御する場合でも基本的には基本色用語(basic color term)が使用されている(全色発話29中16例)。これは黄緑などの色表現が典型性を持たないと操作者が考えることによるものと思われる。現実に少数だが「レモン色」、「黄金色」、「黄緑」などでは操作者の意図すると考えられる制御ができていない。

属性制御において、鮮やかさ(彩度)を表現する適切な言葉が無いために、操作者は「薄い」、「濃い」、「明るい」と言った用語を使用した

れは実験協力者によって適切な解釈が行われず、指示色の彩度方向に関しては最終的な誤差が大きかった。

自由発話群では目標付近になるまでは属性用語がつかわれるが最終的な調整は位置用語で行われている(10人中7人)。

操作者と[機械]のインタラクション：

実際には機械でなく実験協力者が行っているため必ずしも一貫した対応を行っているわけではないが、基本的には[機械]はワードスポッティングで行っている。操作者は、一つの指示用語に対する[機械]の反応を見て、次の用語の選択を行っている。また、[機械]は自己確認の発話(例えば連続した右・右)には一回の発話としてしか対応していない。しかし、本当の機械にこのような特性を持たせるためには発話のリズムやピッチなどの韻律情報を取り入れた分析が必要であることを示している。

程度の副詞は、最後のターゲットに近づくにつれて多用される(属性制御38%、位置制御33%、属性・位置制御23.6%)が、位置制御の場合は[機械]のそれまでの指示用語に対する反応量をベースにして被験者は適宜使い分けている。

程度の副詞が操作開始直後に発話されることは非常に少なく全18名26試行中で2試行でしか見られなかった。普通被験者は「機械」が応答した操作量を見てから程度の副詞を使っている。また、3名の被験者は程度の副詞を使用しても自分の思い通りの操作量とならない場合には、画面上における実距離でのカーソル移動を要求した。

実験1と実験2の結果の1例を図2に示す。

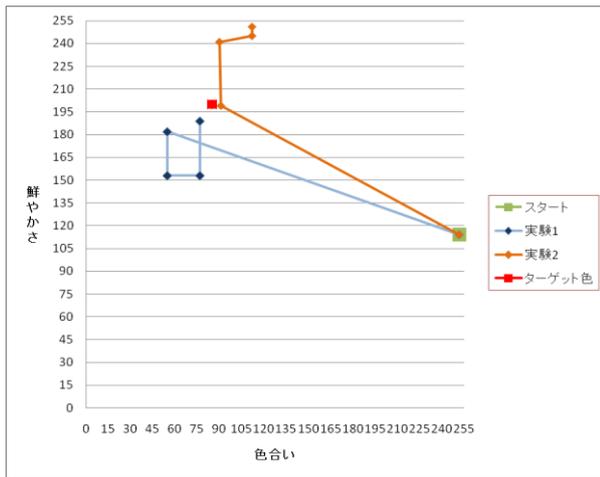


図2 操作結果の例

表1 図2の時の音声発話

1 発話目	黄緑
2 発話目	薄い
3 発話目	右
4 発話目	上

上図は実験1及び実験2でカーソルがたどった軌跡である。ターゲット色に向かっては表1のような発話が行われた。両実験ともに画面右側スタートにカーソルがある状態から音声に従って操作が行われた。一発話毎の動作に合わせたプロットしている。実験2の場合被験者側の適応が起きている場合にはターゲットに近いところへカーソルを誘導出来ているが、実験2のようにこのような適応が無い(オープンループ)場合ターゲットのところへ誘導が上手く行われていない。

実験2及び実験1と実験2の比較から以下のような点が明らかとなった。

属性用語における制御の問題点

実験1においても述べたように色相の表現に関しては日本語において確立している基本色があり、その典型性に頼ってターゲットに近付けることが可能であった。一方彩度に関しては適切に表現する用語が無い場合同一発話に従って

12人の被験者がそれぞれカーソルを制御して得られた最終位置のばらつきは大きかった。

位置制御における特性

位置制御では、前述したように実験2では「機械」の制御特性に対応した発話の調節は無いので、被験者は自信の考える制御量(例えば「右」に対しての移動量)を決めて、それ以降ほとんど変化させることは無いので、被験者各自が考える制御量に従って最終位置はかなりバラツいた。

実験1においては、一見操作者の発話者への適応は無いように見えるが、前述したように発話者の韻律情報の利用が行われる。特に「機械」がカーソルを移動させてから次の制御用発話が行われるまでの時間が短い場合はそれより前の制御が肯定的にとらえられているという判断をしているように見える。

これら一連の実験結果から使いやすい音声制御用の機械には以下のような点が求められる。

自由な音声発話を許容する観点からは、属性制御も位置制御も許容するのが良い。しかし、属性用語を許すと機械側も発話者側と同じ属性認知モデルを持つ必要が生ずる。今回の実験例では、色相に関しては、そのようなモデルを持つことが可能であるが、鮮やかさ(彩度)に関してはそのようなモデルを持たせることは難しい。従ってモデルでは解釈できないような属性用語が使われた場合には機械側からその意味を確認する機能を持たせる必要があると考えられる。また、制御を意図しない自己確認的な発話に対処するために韻律情報を利用した機械側の機能が必要とされるが、今回の実験からだけではデータ量が十分でなくどのような韻律情報を用いれば、このことが可能であるかは解析出来ていない。今後引き続き検討を進めていくことが必要である。