

# 曖昧語を用いた適応型制御システムについて

黄珉錫 橋本周司

早稲田大学院理工学研究科物理学及応用物理学専攻

## 1. はじめに

人はコミュニケーションをとる際に、意図する程度が曖昧な言葉や、話者によりその意味が異なる感性的な表現をよく用いる。特に聞き手、もしくは話し手が専門的な知識を持たない場合は、具体的な数値表現による指示を出すことが難しくなり、曖昧語を用いる以外に自分の意図を伝達することが出来なくなる[1]。本研究ではユーザが曖昧語によって様々な対象を制御することが可能なシステムの構築を目指す。ここでは、音響や画像に対してユーザの主観に基づく指示表現を用いることで、人と対話するように指示できるシステムの試作結果を報告する。

## 2. 方針

### 曖昧語システムについて

「ちょっと」「少し」などの程度を表す表現や「明るく」など質的な方向性を表す表現は言葉の使い手により、また用いる対象により意図が異なる場合がある。本稿ではこれらの表現を曖昧語と呼び、数値表現と合わせてシステムに対して用いる指示表現とする[2]。

### 音響について

はじめに音量、音高についての曖昧語に関する考察を行った。事前実験にて曖昧語に対する認識の傾向を調べ定量化し、その結果を踏まえた曖昧語適応システムの構築を行った。事前実験より、認識に個人差は生じたが各個人には確固とした基準があると考えられた為、ユーザの各曖昧語に対する認識を学習し、適宜修正していくシステムを試作した。

本稿では、ある音響の特定周波数領域を強調もしくは減衰させることによる音の印象を音質と呼ぶ。特定周波数領域をブースト、カットした音響の印象は、音量より顕著に個人差が表れた。同じ音源に対しても個人の受ける印象は多様であり、またその用いる表現も個人により多彩である為、用いる表現を制限するのが困難になる。よって似通っている表現を多く登録、グループ化し個人の好みに合わせて適宜学習できるようにした。同一グループ内の表現間に差異があるときはそれぞれ独立させ学習を続ける。また、音質の場合は音量

や音高と異なり、複数のパラメータを調整する必要がある[3][4]。表現間の関係を把握し適応を行うため、主となる因子を3つとり3次元空間上に各表現を位置づける。各軸には特定の周波数領域を強調、減衰させるフィルタの強さをとる。登録しておく表現の空間上の配置は全被験者の平均値より決定する。ユーザが望む応答とシステムの返す応答が異なったときシステムは更なる指示を受け、ユーザが満足した応答からユーザの認識を学習する。ユーザの求める応答を得ることが困難なときシステムは空間上の近傍からランダムに探索する。

### 画像について

画像処理に関してもまた、指示に対して求められる応答は状況により変わることから、個人差と共に元画像の違いも考慮に入れる必要がある。よってシステム構築は音響の場合と同じ方針とした。各ユーザに対しそれぞれ履歴参照による個人適応を行う。

### マルチシステムとして

本システムはマルチシステムとして、様々なアプリケーションに対して曖昧語指示を可能にする。

## 3. システム概要

システムは以下のように構成した。

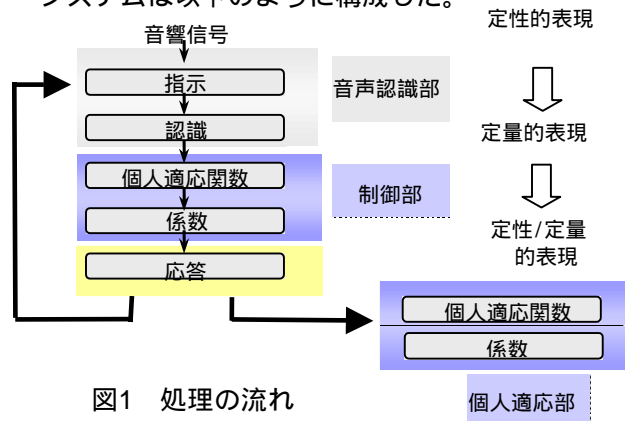


図1 処理の流れ

ユーザの音声入力が行われるとその結果は、ユーザ個人に適した処理を行うためのキー操作に変換される。以下、画像処理過程を中心に述べる。使用したアプリケーションは以下の通りである。

- ・ Microsoft Speech API (Japanese Recognizer v5.1) (音声認識)
- ・ Cycling'74 Max/MSP, Jitter (音響プログラムソフトウェア、制御部、主に音響信号処理)

- Adobe Photoshop7.0 (画像処理)
- Adobe PhotoshopSDK5.0 (画像処理)

## 音声認識

人が意思伝達する際、音声にはその言語情報のみならず、様々な非言語情報が含まれている為、入力装置として、マウスとキーボードのみを用いては本手法の利点が失われてしまう。また、音声で入力することで様々なアプリケーションの並行処理がし易い。よって、より自然なインタラクションをとる為に、音声にてシステムに指示を行うようにした。ユーザはユーザ固有の定性的表現により音声にて指示を行う。認識結果は TCP/IP ネットワークを通じて、MAX/MSP にて構築した制御システムに指示として送信される。音響、画像制御部及び個人適応部はその指示内容から、ユーザに応じた出力を返す。

## 明るさ/暗さ

画像処理を行うケースの一つとして画像の明るさの調整が挙げられる。本稿で用いたソフトウェア、Adobe Photoshop にはレベル補正、明るさ・コントラスト調整、トーンカーブ、明度・彩度や各フィルタなど各種明るさに関する補正手段があり、ユーザによって用いる手段や調整量は異なる。調整にはいくつもの処理を続ける必要があり、複雑な処理を要する。「ちょっと」「少し」「やや」などの程度表現と合わせて行う本システムとの対話学習から元画像やユーザの認識を考慮した処理が可能になる。

## キーアクション

指示を受けたシステムはどのアプリケーションに対する指示であるかを判断する。指定アプリケーションに対応して直接操作、もしくはキー操作として送信する。Photoshop に関しては、指示時に元画像をアクティブにして見ているという前提からキー操作を送信する方式を採った。マウスやキーボードによって多くの操作が必要である処理を、簡易な言語表現で行うことが可能になる。

## 適応

事前実験より、元画像や音響と各指示に対する調整量の間には大きな相関があることが判った [2]。そこで、システムが各ユーザに適応する為に学習機能を用いた。適応過程は次節で述べる。

## 4. 個人適応について

はじめに事前実験のデータを元に学習し、一般的認識を近似した関数を得る。そこからユーザの求める応答と学習関数出力の間に相違が生じたとき、ユーザ固有の偏差として適宜学習し、関数を

修正していく。この個人適応に際しては、調整前の状態と調整量の関係をプロットした不連続点から間を補完し、関数近似をする為にバックプロパゲーション[5]を用いた。ユーザが求める応答が出力されるまで、ユーザの指示とシステムの応答を繰り返す。入出力と中間層の3層の結合関係を考え、出力層に教師信号としてユーザの求めた応答を与える。実際に指示に対して出力された応答と教師信号の違いをもとに結合の修正を行った。追加指示に対する係数は、ある指示が追加の指示であるとき、一回目の指示と比べ、どのような応答をユーザが求めていたのかを最終的にユーザが満足した応答から判断し、適応を行った。

音響の場合の適応例を図2に挙げる。被験者が、学習前後で目的の応答を得るのに要した指示回数を比較した。ここで、学習後とは事前登録した10個の表現に関してそれぞれ8回の学習を行った状態とした。学習前、初期、後期で、被験者10名の平均指示回数の変動を示す。学習前の値は、事前にとった全被験者の平均値とした。

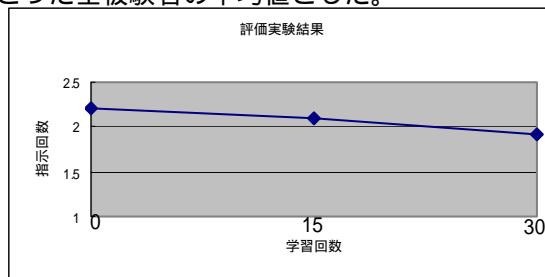


図2 評価実験結果

学習の効果が大きいユーザとそうでないユーザがいたが、平均からも個人適応により求める音響を得るのに要する指示回数が減少することが判る。

## 5. まとめ

計算機との対話において感性的な表現を用いる事でより人間に近い伝達、応答を期待できる。今後、適用したアプリケーションの長期使用により、本システムの効果を研究し、改良を加えるつもりである。

### 参考文献

- [1]竹内伸吾、目黒光彦、金子正秀：“曖昧さを含んだ言葉を用いた空間指示によるコンピュータとのインタラクション” ヒューマンインターフェースシンポジウム 2001、pp.375-378、2001
- [2]黄珉錫、橋本周司：“曖昧語による適応型音響制御システム” 情報処理学会ヒューマンインターフェース研究会研究報告、pp.15-22、2002
- [3]難波精一郎：“音色の測定・評価法とその適用例”、応用技術出版、1992
- [4]中山剛：“音質評価技術の実際(その1)”、日本音響学会誌 44巻4号、1988
- [5]D. Rumelhart and J. McClelland: "Learning internal representation by error propagation", Parallel Distributed Processing 1, 318-362, MIT Press, 1986