

曖昧語による適応型音響制御システム

黄珉錫 橋本周司

早稲田大学工学部応用物理学科

〒169-8555 東京都新宿区大久保 3-4-1

Tel. 03-5286-3233, Fax. 03-3202-7523

{hwang,shuji}@shalab.phys.waseda.ac.jp

あらまし 人はコミュニケーションをとる際に、曖昧な言葉や、話者によりその意味が変わる感性的な表現をよく用いる。曖昧な表現を数値化することは困難であるが、ニュアンスを伝える手段として人間と機械のより円滑なコミュニケーションを実現する為に欠かせないものである。本研究では対象を音響に限り、ユーザの曖昧語による音響制御が可能なシステムの構築を目指している。ユーザは定性的な表現で指示することにより、音響制御に関する知識を必要とせず、意図する音響を得ることができる。ここでは曖昧語表現の定量化と個人性を検討する為の予備実験の結果と適応型音響制御インターフェースの試作について報告する。

Adaptive Sound Control System Using Ambiguous Expressions

Min Sok Hwang Shuji Hashimoto

Dept. of Applied Physics, Waseda University

3-4-1 Okubo, Shinjuku-ku, 169-8555 Tokyo

Tel. 03-5286-3233, Fax. 03-3202-7523

{hwang,shuji}@shalab.phys.waseda.ac.jp

Abstract In verbal communication, we often use the expressions whose meanings or intentions are ambiguous. We consider that, such kinds of expressions are also important in order to establish smooth communication between human and computer, although it is difficult to adopt them quantitatively. In this paper, we focus on sound control system using ambiguous expressions. First, we show some preliminary experimental results to examine the possibility for the numerical evaluation of ambiguous expressions and to characterize the personal dependency of the expressions. Next, an experimental sound control interface is shown. This system enables us to control sound using qualitative expressions without demanding special acoustical knowledge.

1. はじめに

現在コンピュータは技術者や専門家のみならず、多くの人に用いられているが、計算機には一般的に曖昧語の理解が困難で、ユーザの定量的な指示を仰ぐ必要がある。コンピュータと相互の情報伝達を行う上で、感性的な表現を用いる事によって、より人間に近い伝達、応答が可能で、ユーザ側にも新しい発見が期待できる[1][2][3][4]。

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えることが出来なくなる[5][6][7]。この場合例えば、「もう少し」という言葉が与えられただけでは、どのくらいの量や幅であるのか判断できないが、その言葉の話者が誰であるか、また、状況も考えた上で話者の意図を推測できることがある。

ここで、曖昧語とは「少し」、「ちょっと」等、意図する程度が曖昧な表現や、「明るく」、「透明感がある」等、意図が定量的に定義し難い感性的な表現を含めたものとする[8]。これらは聞き手にニュアンスを伝える手段として意思伝達に欠かせないものである。

本研究ではコンピュータとのコミュニケーションにおいて曖昧語を用いることを目指し、特に対象を我々にとって身近な音響に特化して実験を行った。はじめに各ユーザと各曖昧語の関係を調べ、定量化する為に事前実験を行い、その結果を踏まえて曖昧語適応システムの構築を行ったので、それらについて報告する。

2. 事前実験

2.1 実験手法

まず、曖昧語が人によりどう認識されているのかを調べる為、事前実験を行った。ここでは最も基本的な音量について考え、「ちょっと大きく」、「もう少し小さく」などの指示が与えられたとき、人がどれくらいの応答を返すのかを見ることで

各々の表現の捉え方の違いを調べた。用いる表現は限定し、実験は無音時、騒音レベルが48dBの研究室で行った。音源は正弦波及び音声、音楽などとし、被験者は大学生及び大学院生25名とした。実験ではグラフィカルなインターフェースを用いることで、被験者が数値を意識せずに聴覚のみをたよりに音量を調整するようにした。

事前実験 1

はじめに単純な波形を音源として、一般的な認識の傾向を調べた。コンピュータ上で、正弦波の足し合わせによる発振器を作成し、データの取得を行った。本実験では、人間にとって不快に感じない周波数として、基本周波数を950Hzとした。また、基準音の音量や周波数を変えて個人の曖昧語の認識との関連を調べると共に、複数の被験者を対象とし、同一条件で実験を行うことにより個人差も調べた。

曖昧語として、以下の表現を用いた。

ちょっと大きく	ちょっと小さく
少し大きく	少し小さく
気持ち大きく	気持ち小さく
大きく	小さく

作成した発振器は、基準音とその8倍音までの正弦波の足しあわせにより音を出すものである。基本周波数と波形は可変で、被験者は音を「ちょっと、少し」/「大きく、小さく」等、指示に従って音量を調整する。音量はキーボードの矢印キーを押し続けることにより滑らかに変えることができる。実験では、システムの「音を大きくしてください」という指示に対し被験者がどれほどの応答を返したのかを調べた。被験者は現在の音量と調整量を視覚によらずに判断し、各指示に対し自分の満足するまで調整を行う。設問と応答のログを自動的に記録し、検証した。

事前実験 2

次の実験は音楽プログラミングソフトウェア

MAX/MSP を用いて作成したパッチ上において行った。MAX/MSP を用いることでより音響に対しより多彩な制御を感覚的に行うことができる[9]。被験者は、実験 1 と同様にグラフィカルなインターフェースを用いて音量を調整するようにした。

システムからの指示には事前実験 1 と同じ表現に、「やや大きく/小さく」を加えたものを用い、発振器、もしくは音楽を音源として、元音量のレベルを変えながらそれに伴う応答の変動を調べる。

2.2 事前実験結果

事前実験 1 より、以下の結果が得られた。図 2.1 は「ちょっと大きく/小さく」という指示を出したときに被験者が調整した音量を示したものである。横軸には各サンプルの番号、縦軸には調整した音量をとっている。グラフでは各被験者二回の試行の結果を隣り合わせてプロットしている。図 2.2、図 2.3 は「少し大きく/小さく」、「大きく/小さく」の結果を示している。

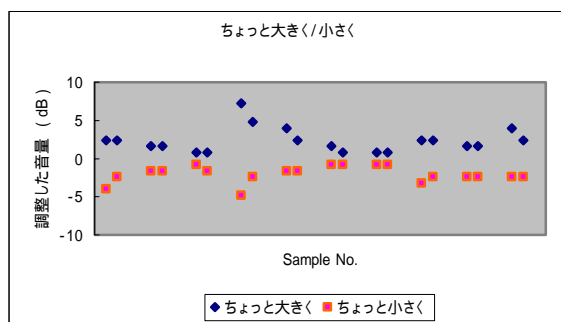


図 2.1 被験者による調整量の違い(ちょっと大きく/小さく)

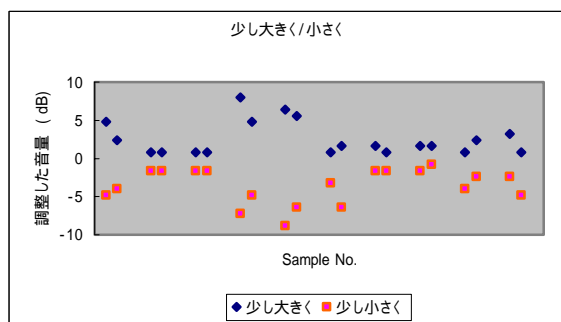


図 2.2 被験者による調整量の違い(少し大きく/小さく)

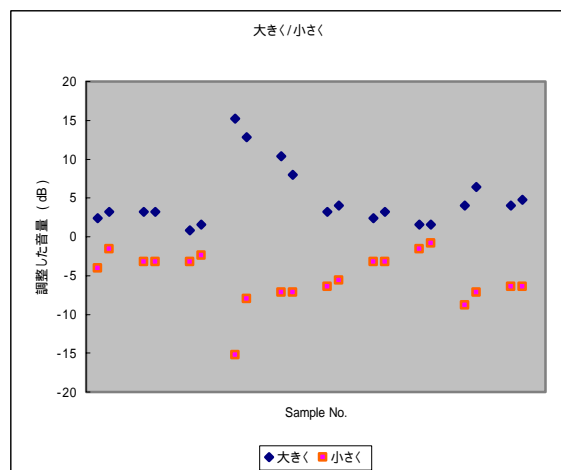


図 2.3 被験者による調整量の違い(大きく/小さく)

同じ条件で、同じ指示を与えても被験者によって調整量は大きく異なっていることがわかる。しかし、各被験者の二回行った試行の調整量を比べると似通っていることから、曖昧語に対する認識は、被験者ごとにはばらつきがあるが、各個人はほぼ一定であると考えられる。また、0dB を境界としてグラフを折り返すと、ほぼ上下の点が同じところにプロットされる。ここから、程度を表す副詞が同じだと、大きくするときと小さくするときで調整量は近くなると考えられる。「大きく/小さく」に関しては限定する副詞表現がないため個人差がより大きく出ているが、ここでも同一被験者の 2 回の試行での変動は少ないといえる。

また、周波数を変えたとき、被験者の返す応答に変化がでた。図 2.4 は正弦波を音源として、52dB の元音量で「ちょっと大きく」という指示を与えたとき、被験者の応答の平均を示す。

周波数	600Hz	1kHz	4kHz	6kHz
調整量	5.6dB	7.2dB	4.0dB	6.6dB

図 2.4 周波数と調整量の関係

得られた結果より、周波数による調整量の変動は等ラウドネス曲線(図 2.5)を元に近似できると推測される。

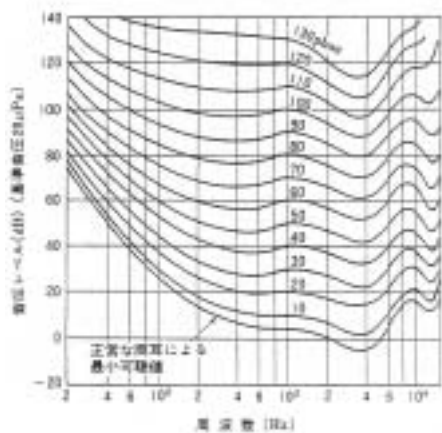


図 2.5 ロビンソン、ダッドソンによる等ラウドネス曲線[10]

つまり、音源から基本周波数を求め、調整量を変えることで、よりユーザの求める応答を高い精度で返せることが期待できる。

次に事前実験 2 の結果について示す。本実験ではまず、元音量を変えることによる調整量の変化を調べた。ここでは「ちょっと大きく」という指示を例に挙げ、元音量が 65dB のときと 75dB のときを比較した結果を図 2.6 に示す。

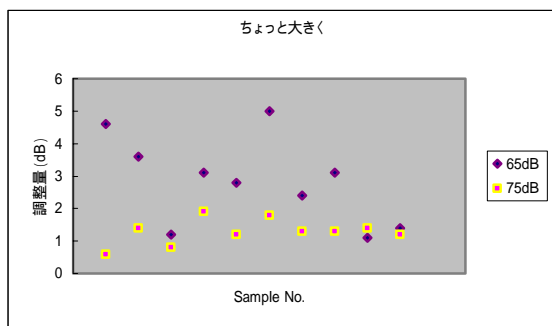


図 2.6 元音量の違いによる調整量変動

横軸に各被験者を、縦軸に調整量をとった。与えられた指示が同じでも、その元音量に対し、被験者が既に大きいと感じている場合、調整量が小さくなっていることがわかる。「小さく」という指示に対する応答にも同じことが言えた。また個人差も大きく出ていることより、元音量と各指示に対する調整量の間にある相関関係を調べ、ユーザ固有に適応していく必要があるといえる。

次に表現間の調整量の関係を比にとって示した。まず「少し大きく」と「ちょっと大きく」という表現で、同じ条件のもと調整量はどう異なる

かを調べた。

$$\frac{\text{少し大きく}}{\text{ちょっと大きく}} = 0.834 \quad (1)$$

このように、一般に「ちょっと」の調整幅を大きくする傾向があった。平均的には「少し」と「ちょっと」の関係はこのようなになったが、大小関係を逆にとる人もいた。

$$\frac{\text{ちょっと小さく}}{\text{ちょっと大きく}} = 1.007 \quad (2)$$

次に元音量が 65dB のときにおける「ちょっと大きく」と「ちょっと小さく」の関係をみると、ほぼ同じ調整量となった。しかし、元音量によってはこの関係は変化する。

また、「ちょっと大きく」、「もうちょっと大きく」と、繰り返し指示を出したときの調整量の変動を調べたところ、後の指示になるにつれて調整量が大きくなる人と、小さくなる人に二分された。しかし、前者に属した人は、常に後の表現になるにつれて調整量を大きくする傾向があった。後者に関しても同様に一貫性があった。

3. 試作システム

3.1 方針

事前実験の結果をまとめると以下の様になる。

- (1) 「大きく」という指示を与えたとき、元音量が非常に小さいときや、被験者が元の音が既に十分大きいと感じると、指示に対する調整量は減少する。
- (2) 各々の被験者は元の音が同じであれば、ある指示に対して毎回ほぼ同じ応答を返す。
- (3) 「ちょっと大きく」の後「もうちょっと大きく」と繰り返し指示を出したとき、後者になるにつれて調整量が大きくなる人と小さくなる人に二分される。
- (4) 「ちょっと」と「少し」等の表現は人により大小関係が逆転するが、ある一人に関しては一定である。よって、一

一般的傾向にとらわれない個人適応が必要である。

以上より、システム設計の方針を以下のように定めた。まず、元音量と調整量の間に関連関係が見られた為、以下の図のような調整関数を導入する。

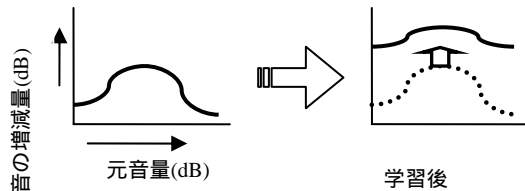


図 3.1 ユーザ A の「もうちょっと大きく」の調整関数

調整関数は学習によってユーザに適応するように変化するが、その初期値は事前実験結果の平均値とした。また、追加指示に対する係数を導入し、この係数も学習により求めることにした。

3.2 試作システム概要

事前実験より、曖昧語に対する認識に個人差はあったが、各個人は一定の基準をもっていると考えられる。よって、システムが各個人の基準を学習することにより、ユーザの求める応答に近づけることが期待できる。システムはユーザの指示と、応答に対するユーザの反応を履歴として残し、そのデータを用いて個人適応学習を行う。曖昧語を用いて、被験者に指示を繰り返させることで各表現に対する変化量を調べ、そのデータを逐次システムに蓄積、学習させた。

デフォルトでは曖昧語として事前実験で用いた表現が使えるが、他の表現も学習により受け入れることができるようにした。音源は aiff ファイル、MIDI ファイルもしくは発振器より選択できる。ユーザは上記の曖昧語による指示や、直接数値を変えることで各パラメータを定性的または必要に応じて、定量的に調整することができる。実験に備えて、音量、調整幅などはユーザに見えないように設定した。与えられた指示は、データ処理、蓄積された後応答として返される。

ユーザが満足できる応答を返されたとき、システムに OK signal を送ると、システムはそれまでの履歴から対象ユーザの曖昧語の認識を推定し、データベースに蓄積し以後の応答に反映させる。

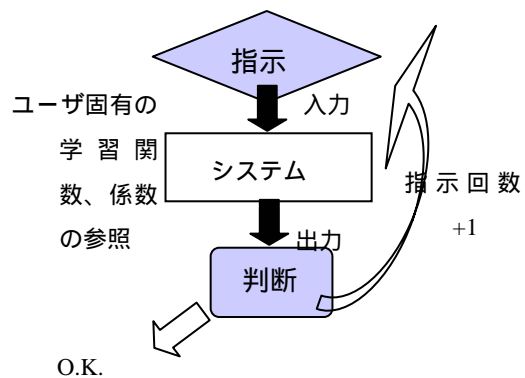


図 3.2 入出力の流れ

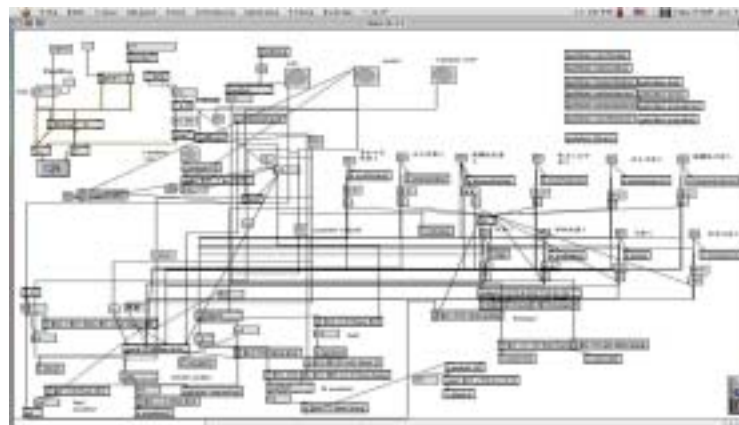


図 3.3 試作システム

3.3 学習部

はじめに事前実験のデータの平均から調整関数の初期値を決め、そこからユーザの求める応答と調整関数の間に相違が生じたとき、調整関数を適宜学習、修正するようにした。調整関数は3層パーセプトロンで計算することにして、学習にはバックプロパゲーションを用いた。学習過程ではユーザの指示と、システムの応答を、ユーザが求める応答が出力されるまで繰り返している(図3.2)。つまり、入出力層と中間層の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、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。

3.4 音声認識部

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。そのことを考えた上でも、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。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。音声認識エンジンとしては、Microsoft Japanese ASR Version 5 Engine 及び Microsoft English Recognizer v5.1 を用いて、pc上で得た認識結果を MAX/MSP へ LAN を通し送信するシステムを構築した。

図3.4は、音声入力システムの処理について示したものである。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。システムは認識結果を TCP/IP ネットワークを通じて、MAX/MSP へ指示として送信する。音響制御部及び個人適応部はその指示内容から、指示を与えたユーザに応じた出力を返す。

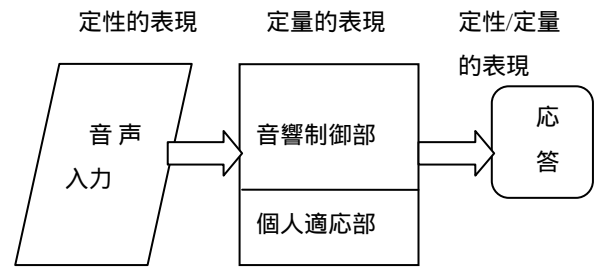


図3.4 処理の流れ

本稿で試作したシステムにおいては、認識対象となる言葉の数はさほど多くないので、高精度の認識結果を得ることができた。音声入力を導入することで、ハンズフリーで操作できる利便性が得られるばかりでなく、より自然にシステムを使えるようになった。また、手などの操作を必要とするユーザインターフェースよりも評価実験の精度が上がるのが期待できる。

3.5 評価

前節で試作したシステムの評価を行った。図3.5、図3.6は「ちょっと大きく」、「少し小さく」と指示を出したときの調整量の変動過程である。2人の被験者が行ったそれぞれの指示に対して学習を重ね、各表現に対して学習前、5回、10回の調整量の変動を調べた。ここで学習回数とはユーザが満足する出力を得た回数、つまり適応を行った回数とする。音源は周波数 950Hz の正弦波、指示前の基本音量は 55dB である。

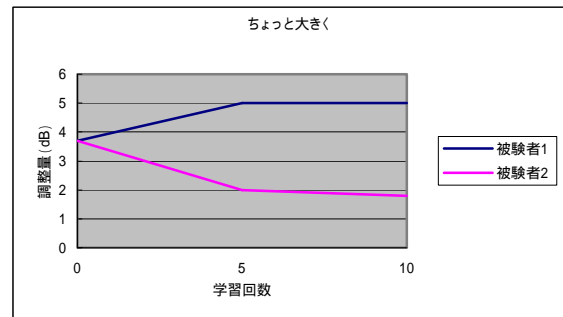


図3.5 「ちょっと大きく」の個人適応過程

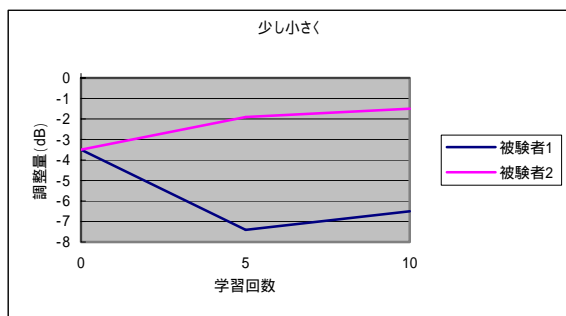


図 3.6 「少し小さく」の個人適応過程
学習を 10 回行ったところで、ほぼ収束しており、ユーザがシステムの返す応答に満足していることがわかる。

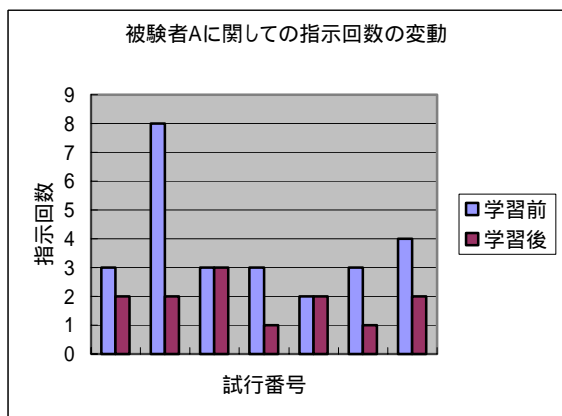


図 3.7 学習前後の指示回数
図 3.7 は、ある被験者が 7 回のテストを行い、学習前と後で目的の応答を得るのに要した指示回数を比較したものである。ここで、学習後とは全ての表現に関して 8 回以上学習を行った状態とした。グラフより、学習後の方が少ない指示回数で目的の応答を得ていることがわかる。特に、2 回目のテストに関して学習前には 8 回の指示を要した。これは被験者の意図と、用いられた表現の調整関数の間に隔たりがあった為である。学習後には少ない指示回数で満足する応答を得ていることから、学習を行うことでシステムは被験者に適応しているといえる。

次に事前実験 2 で示した 10 個の曖昧語指示に対し、ランダムに提示した条件下で学習した際の評価実験の結果を示す。図 3.8 は同一の条件で、求める応答を得るまでに要する指示の回数の変動を表している。ここで指示回数は被験者 7 名の

平均である。

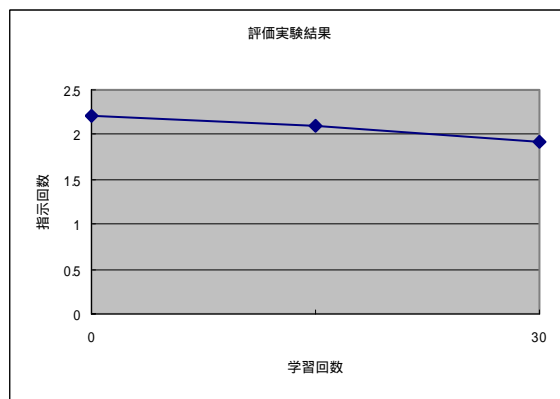


図 3.8 評価実験結果
学習する表現の数が多いと、短い学習回数では成果をあげることは難しい。しかし、この結果より滑らかな傾きではあるが、学習することにより指示回数を減らすことができることがわかる。

4. 音量以外の要素について

4.1 音高

音高についても音量と同じように制御可能とした。音量に比べると音高は楽曲のキーの調整など用途は限られるが、有効な調整手段となる。しかし、音高は一般に音量に比べて身近でないので個人差が大きく出るとされる。よって、個人適応は有効であるが、個人の認識自体も不安定である。また、聴覚上、音の強さとの関係も含めて考慮する必要があるだろう。

4.2 音質

音質（音の印象）については、個人差がより現れやすく、本手法は有効と思われる[11][12]。「やわらかい音にする」という指示があっても、音響の知識がないと制御できず、話し手自身も曖昧語を用いないと表現できない場合も多い。また、知識があっても話し手の意図を推測するのは困難である。音量、音高はひとつのパラメータで制御可能であるが、音質の場合は、複数のパラメータを調整する必要がある。しかし、調整法としては、音量や音高と同じであり、現在、音質が「やわらかいーかたい」という表現に限定してフィルター

のカットオフ周波数を制御することをここで述べた手法を適用して検討しているところである。

5. まとめ

本稿では、曖昧語を用いて音響制御を行うシステムを構築した。本システムを用いることでユーザの求める応答を定性的な表現から得ることができた。また、特に音質に関しては適応方法を考慮することで更に融通の利くシステムにすることができるであろう。本システムを用いることで、ユーザ固有の表現であってもその意図をシステムが学びとることができる。曖昧語適応システムは今後、音響に限らず様々なインターフェースに応用することが期待できる。

参考文献

- [1]Perry R Cook: “Music, Cognition, and Computerized Sound”, The MIT Press, 1999
- [2]大照完、橋本周司:“ 仮想音楽空間 ”、オーム社、1994
- [3]長嶋洋一、橋本周司、平賀譲、平田圭二 :“ コンピュータと音楽の世界 ”、共立出版、1998
- [4]小長谷尚紀、橋本周司 : “ 音響信号の時系列分析に基づく音楽の自動分析の試み ”、情報処理学会全国大会講演集2、p.p.271-272、2001
- [5]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
- [6]倉橋哲朗、國分三輝、古西浩之、向江秀之、井口弘和、川澄未来子 :“ ニューラルネットワークを応用した感情理解型設計支援システム ”、信学技報 HIP2000-56(2001-3)、pp.9-14
- [7]新山祐介、徳永健伸、田中穂積 :“ 自然言語を理解するソフトウェアロボット：傀儡 ”、情報処理学会論文誌(2001-6)、pp.1359-1367
- [8]黄珉錫、橋本周司 :“ 曖昧語による音響制御について ” 電子情報通信学会全国大会、 pp. 310、2002
- [9]ノイマンピアノ(赤松正行、左近田展康):“ トランス Max エクスプレス ”、リットーミュージック、2001

[10]橘秀樹 :“ 建築音響(基礎)・教材 ”、東京大学生産技術研究所、1999

[11]B.C.J.ムーア :“ 聴覚心理学概論 ”、誠信書房、1994

[12]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