

## 実世界のもののアフォーダンスを利用したインタフェース における音声利用方法の検討

伊賀 聡一郎<sup>†</sup>, 安村 通晃<sup>†</sup>

慶應義塾大学大学院政策・メディア研究科<sup>†</sup>

〒 252 神奈川県藤沢市遠藤 5322

igaiga@sfc.keio.ac.jp

<http://www.sfc.keio.ac.jp/~igaiga/>

### Abstract

我々は実世界のものからのアフォーダンスの計算機インタフェースへの応用を試みている。本研究では、実世界のものを利用した計算機システムにおける合成音声のアフォーダンスに関する実験を通じて実環境の音声情報による拡張に関する検討を行なう。合成音声に対する知覚の際のカテゴリーの判断基準、SD法による合成器のパラメータ変化による印象の違い、実環境の指示操作における合成音声の影響に関しての実験を行なった。合成音声のカテゴリーの種類による実操作でのモードの表現、合成器のパラメータの切替えによる実環境の状況表現の可能性について検討した。また、実世界の指示操作でのパラメータの変化による影響について実験を行ない、心理的な印象を元にした合成器のパラメータの切替えについての可能性を示した。

## Applying Speech Synthesizer using Real World Affordance in Computer Interface

Soichiro Iga<sup>†</sup>, Michiaki Yasumura<sup>†</sup>

Graduate School of Media and Governance, Keio University<sup>†</sup>

### Abstract

This paper describes on applying speech synthesizer in computer interface which utilizes real world affordance. We have conducted three experiments which are to categorize voices of speech synthesizer, define mental images of voices by the difference of parameters (accent, pitch, and speed), and observe the actions of users in instructed tasks. And we have shown the possibility of using speech parameters in the computer augmented environment by representing modes by changing parameters of speech synthesizer.

## 1 はじめに

マイクロコンピュータの発達によりハイテク製品が身のまわりにあふれており、小型化や携帯化が進みながらも複雑な機能を実現している。しかし、ビデオ機器の録画操作ひとつをとってみても、必ずしも誰もが利用できるインタフェースが提供できているとはいえない。この場合、ボタンを押すという人間の行為とシステムが録画を行なうという実際の動作との間には常に獲得された「知識」が要求され、利用における障壁となっている [7]。

人間は様々な情報を提供する環境に取り囲まれ、環境の提供する情報から行為を発見する。これら環境にある情報はアフォーダンスと呼ばれ、我々は行動に際し潜在的に環境からアフォーダンスを知覚している [4, 8]。計算機システムやハイテク製品の増加により、これらに増幅された実世界も我々がアフォーダンスを感じるひとつの新しい環境として認識していく必要がある。

我々は実世界の「もの」から感じるアフォーダンスを利用した計算機インタフェースの試みを行っている [5, 6]。実世界のものを計算機や電子機器のインタフェースとして利用することで、人間にとってより直接的な知覚が可能なシステムが期待される。計算機により拡張された世界を含めた「環境」の中から人間が何を感じ、行動を起こし、その行動に対して計算機が適切なフィードバックをいかに行なうかが重要な課題である。

本稿では、実環境のアフォーダンスを利用したインタフェースで合成音声を利用するにあたっての基礎的な方法について検討する。合成音声を知覚するにあたってのカテゴリ実験、合成音声器のパラメータ変化による認知実験、実環境の指示操作における合成音声のパラメータの違いによるユーザの行動の変化に関する実験の3つの実験を行なう。これらの実験を通じて、実世界を拡張・増幅するひとつの手法としての音声による情報提示方法に関する検討を行なう。

## 2 実世界のアフォーダンスを利用したインタフェース

我々は、実世界アフォーダンスを利用したインタフェースの可能性を示すため、プロトタイプシステムの実装を行なっている [5, 6]。本システムでは、実世界のものの操作や複数のものの関係により計算機や周辺機器の操作を行なうことができる。これにより、実世界のものからのアフォーダンスを利用したインタフェースが可能になる。

例えば、箱は「入れる・しまう」ことをアフォードすることを利用して、箱に何かを入れることをシステムへのトリガーとする、といったことが考えられる。

図1にシステムの構成を示す。机上のものはWSに接続されたビデオカメラによりとらえられ、ものの種類と位置が認識される。現状では、ものにそれぞれユニークな色を割り振ることで個々のものを認識している。検出されたものの情報は、スクリプト解釈部に送られ、ものの状況や関係が決定される。これらのものの関係から、フィードバックの選択を行ない、実際の機器の操作や合成音声による情報提示が行なわれる。

## 3 拡張された実環境における合成音声のアフォーダンス

計算機により拡張 (Augment) された環境を新たな実環境としてとらえた場合、計算機からの補助的な情報提示として合成音声を用いることが考えられる。ここでは、3つの実験を通じて、新しい拡張された実環境での合成音声の利用に関する検討を行なう。まず、知覚カテゴリ実験では、人間が様々な合成音声をどのような判断基準で分類するかについて検討する。次に、SD法を用いた合成音声の印象の違いについて実験を行なう。さらに、実世界の指示操作における合成音声の影響について分析を行なう。

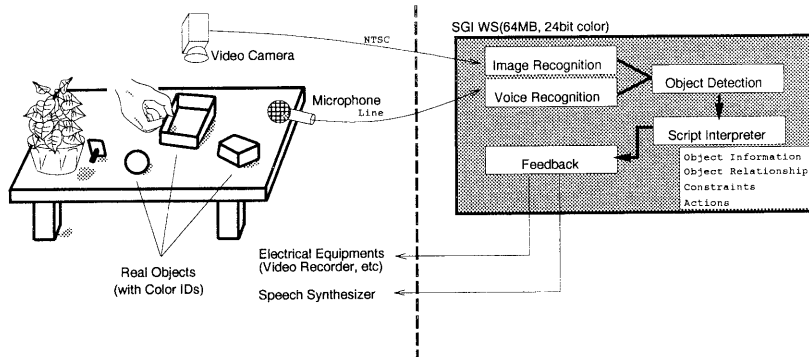


図 1: システム構成

### 3.1 合成音声のカテゴリ実験

り化され、キーワードが抽出された。

#### 3.1.1 目的

既存の合成音声を知覚するときの刺激に対するカテゴリー化の判断基準を明確にすることで、システムの目的に合った合成音声の提示が可能になる。そこで、様々な種類の既存の合成音声を提示した際の、ユーザの分類の判断基準とカテゴリーの数や種類に関する実験を行なう。

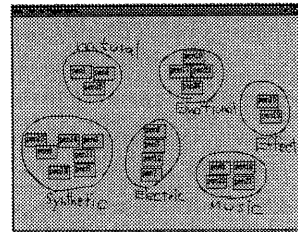


図 2: グルーピング結果画面例

#### 3.1.2 方法

Macintosh 上における 24 種類の合成音 (通常の合成音声、エフェクト処理、音階付きなど) を自由に分類してもらう。分類後に分類されたそれぞれのグループに被験者が分類の判断基準としたキーワードをつけてもらう。

カテゴリー化されたグループの平均グループ数は 5.4 であり、最大 8 グループ、最小 4 グループであった。キーワードとして出現した単語の総数は 32 種類であり、代表的なものは表 1 のようにまとめられる。

被験者はウィンドウ上のアイコンの操作で個々の合成音声ファイルの操作を行なう。アイコンをクリックすることで合成音声 ("This is a pen") が提示され、アイコンのスクリーン上での移動により空間的な分類を行なう。

表 1: キーワードとして利用された単語

#### 3.1.3 被験者

被験者は、大学生・大学院生 11 名である。

性別に関する単語	男性・女性・重い・軽い・高い・低い など
年齢に関する単語	子供・老人・おばあさん など
合成の質に関する単語	人間的・機械的・標準・自然・合成的・まとも など
加工処理に関する単語	音楽・電子的・エフェクト・フィルタ・笑い など

#### 3.1.4 結果

図 2 はグルーピングの結果の画面例である。被験者の自由なアイコン操作により刺激音声がかテゴ

抽出されたキーワードを提示した刺激音別に集計を行ない、最も出現頻度の高かったキーワードで意味付けを行なったところ、「標準」、「効果音」、

「電子的・合成」,「音楽」,「感情」といったキーワードで刺激音を分類することができた。

## 3.2 合成音声のパラメータが与える印象に関する実験

### 3.2.1 目的

既存の音声合成装置には、アクセント・ピッチ・スピードといったパラメータがある。ここでは、システムが情報提示する際の、これらの合成音声のパラメータとユーザに与える印象の関わりを明らかにする。実験にはSD(Semantic Differential)法を用いて、合成音声のパラメータの違いとユーザの受ける感覚概念の評価を因子分析を通じて行なう。

### 3.2.2 方法

合成音声には、Macintosh上で動作する音声合成ソフトウェア(Ricoh製雄弁家ver2.0)を用いた。ソフトウェア上でアクセント・ピッチ・スピードの3つのパラメータの様々な組み合わせにより刺激となる音声を生成する。これらのパラメータの組み合わせから18種類の刺激の提示を行なう。刺激の提示はランダムに行なわれる。被験者には5段階評価の12の評価語(形容語対)が計算機のウィンドウ上に提示され、一定時間(約30秒)内にすべての項目に関しての選択を行なってもらう。評価語は12種類用意した(表2参照)。

表 2: 評価語一覧

軽い(light)	-	重い(heavy)
親しみにくい(unfamiliar)	-	親しみやすい(familiar)
積極的な(active)	-	消極的な(passive)
きりっとした(smart)	-	ぼんやりとした(blockhead)
のどかな(calm)	-	緊張感に満ちた(tense)
女性的な(womanly)	-	男性的な(manly)
乱暴な(rough)	-	ていねいな(polite)
きらい(dislike)	-	好き(like)
明るい(sunny)	-	暗い(gloomy)
不愉快な(unpleasant)	-	愉快な(pleasant)
静かな(quiet)	-	にぎやかな(lively)
みにくい(ugly)	-	美しい(beautiful)

### 3.2.3 被験者

被験者は、大学生・大学院生30名である。

### 3.2.4 結果

評定データを主因子法により因子分析し、ヴァリマックス法により回転した。その結果、「性別に関する因子(男性的・女性的)」、「親近性に関する因子(親しみやすい・親しみにくい)」、「情緒・評価性に関する因子(緊張感・のどか、にぎやか・静か)」の3因子が得られた。それぞれのパラメータの因子得点と得られた因子との関連から、アクセントの強弱が「親近性」、ピッチの高低が「性別」、スピードが「情緒・評価」のそれぞれの因子を表す傾向が観察された。

図3は合成音声のパラメータ特徴別の評定データの平均と評価語との関係を表している。横軸の1と5がそれぞれ「強くそう思う」、3が「どちらでもない」、2と4がそれぞれの中間を表す。また、データはアクセント強(AS)、アクセント弱(AW)、ピッチ高(PH)、ピッチ低(PL)、スピード速(SF)、スピード遅(SS)を表しており、18種類の刺激のうち、それぞれのパラメータが最大あるいは最小のものを平均している。

図3のそれぞれの特徴的なデータの平均値においてもパラメータと3因子の関連において同一の傾向が見られる。例えば、男性的(manly)・女性的(womanly)の評価語対では、ピッチ低(PL)が「男性的」をピッチ高(PH)が「女性的」を表しているように、それぞれのパラメータにおいても、同様に因子との関わりが観察された。

## 3.3 実環境での指示操作におけるユーザの反応に関する実験

### 3.3.1 目的

前述の合成音声のパラメータ実験から得られた因子では、親近性と情緒・評価性の2つの因子がユーザの行動への影響が考えられる。その中でも因子得点が特徴的であった4つのパラメータの組み合わせを利用して、実際のものの操作へ適用した。ここでは、パラメータの違いが、ユーザの行動に及ぼす影響に関しての実験を行なう。

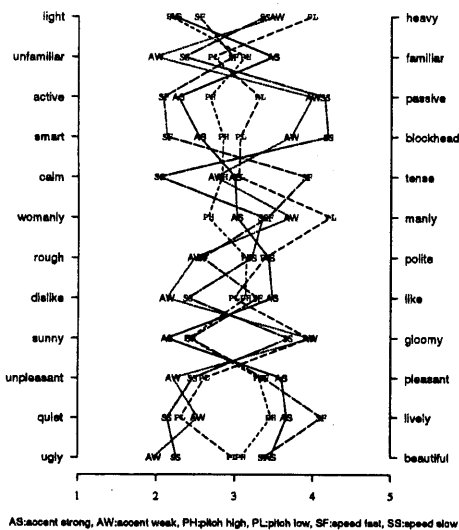


図 3: パラメータ別の平均得点

### 3.3.2 方法

表 3 の合成音声ソフトウェアのパラメータの組み合わせにより音声を合成した。これらの合成音声によりユーザにものの操作を行なうよう(“赤いボールを緑の箱に入れて下さい”など)指示した。被験者のもの操作に関して、ものに手が接触してから非接触の状態になるまでの時間を計測した。

表 3: 因子とパラメータの組み合わせ

因子	アクセント	ピッチ	スピード
親しみにくい・のどか	Weak	Low	Slow
親しみにくい・緊張感	Weak	Normal	Fast
親しみやすい・緊張感	Strong	Normal	Fast
親しみやすい・のどか	Strong	High	Normal

### 3.3.3 被験者

被験者は、大学生・大学院生 9 名である。

### 3.3.4 結果

図 4 に実験の結果を示す。R と B はボールの種類(赤・青)、I と O はそれぞれボールの操作(入れる・出す)を表す。全体としては、パラメータ(Weak, Low, Slow)でもものに接触している時間が長く、その他では比較して短い傾向がみられた。パラメータはアクセント、ピッチ、スピードの順である。t=5%で統計的検定を行なったところ、入れる・出すの双方に関して、パラメータ(Weak, Low, Slow)と(Weak, Normal, Fast) (Strong, Normal, Fast)などに有意差がみられた。また、BI と BO に関しては、操作の違いに関してにも有意差があった。

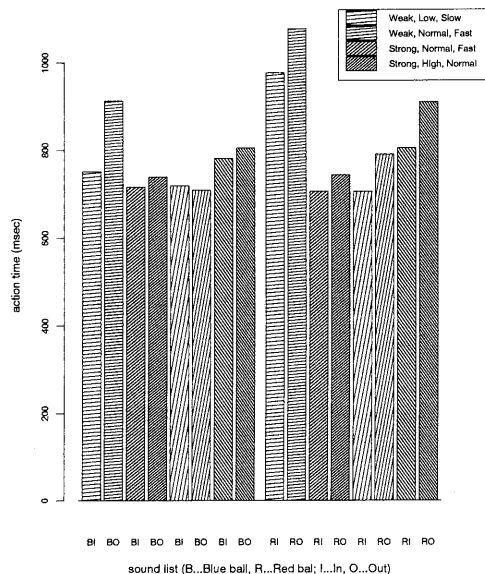


図 4: 音声ファイル別のものの操作に費やした時間

## 4 考察

合成音声の分類実験では、様々な合成音声が表示された際のカテゴリ別の知覚基準として、「標準」、「効果音」、「電子的・合成」、「音楽」、「感情」といったキーワードが得られた。システムを持つ「パーソナリティ」を除いた場合、個々の音声は

これらの判断基準においてユーザに知覚されることが考えられる。本実験の結果から、実環境を拡張するインタフェースに音声を用いる際に、機器の各モードを表現するのに利用することが考えられる。

合成音声のパラメータが与える印象に関する実験では、因子分析を行なった結果、「性別に関する因子」、「親近性に関する因子」、「情緒・評価性に関する因子」の3因子が得られ、ピッチは性別、アクセントは親しみやすさ、スピードは情緒・評価性に関連がみられた。ある合成音声による情報提示を行なう際に、これらのパラメータを変化させることで、ユーザへのシステムの心的な位置付けが可能になる。

実環境の操作でのユーザの反応に関する実験では、実際に様々に変化させたパラメータによる合成音声による作業の提示を行なった。その結果、スピードが遅い場合に操作時間が長くなる傾向がみられた。また、統計的な有意差はみられなかったが、情緒・評価性の因子（緊張感・のどか）において特徴的であったデータ（Strong, High, Normal）では、スピードが標準的であるにも関わらず、作業時間が長くなる傾向がみられた。これは、スピードが速い場合に操作も速くなるという他のデータの傾向に反しており、情緒性に関する因子が影響していることが予想される。

## 5 問題点と今後の展開

システムがあるパーソナリティを提供する際には、効果的な合成音声の区別よりも、エージェント型システム [2] がとる手法のように、性別やピッチの変化が影響を及ぼすと考えられる。また、実環境との結び付きを考えた場合、単に知覚できる音声の種類を切替えるのみではなく、実際に操作しているものに関連した音声提示が必要であろう。実世界のものと同様に関連性については今後の課題としたい。

合成音声のパラメータが与える印象に関する実験で提示した音声は、アクセント、ピッチ、スピードの標準的な合成器のパラメータを変化させている。しかし、実際に拡張現実感システムにおいて、実世界の様々な状況に結び付いた合成音声による状況の変化をユーザに知覚させるにはより多くのパ

ラメータが関連してくることが考えられる。また、単に音響的な変化だけではなく、状況や意図といったコンテンツとの関連も考慮する必要がある。

実環境の操作でのタスクは、2つのボールの箱への出し入れという比較的容易なものであった。これにより実験中のタスクに関しての学習効果が発生してしまうおそれがある。さらにタスクを複雑にして検証する必要がある。実際の利用場面においては、合成音により単調な作業の意欲を持続させることが報告されている [1]。これらについても今後の研究課題としたい。

## 6 まとめ

実世界のものによる操作によるインタフェースでの合成音声の利用手法に関する検討を行なった。人間の合成音声に対する知覚のカテゴリー基準、パラメータ変化による印象の違い、実世界での指示操作における観察を通じて、機器のモードの表現やユーザの実際の操作への合成音声の応用について検討した。

新しい実世界の環境として拡張現実感をとらえることで、計算機システムやハイテク機器との容易なインタラクションが可能になると考える。

最後に、実験に御協力頂いた慶應義塾大学大学院政策・メディア研究科マルチモーダルインタフェースプロジェクトの皆さんに感謝申し上げます。

## 参考文献

- [1] 阿部匡伸, 誤り指示音声の特徴分析と音声出力への適用, 電子情報通信学会論文誌 D-II Vol. J79-D-II, pp.2191-2198, 1996.
- [2] 安藤ハル, 畑岡信夫, 対話機能を有するエージェント型インタフェースのユーザ学習効率についての評価, 日本音響学会講演論文集, pp.191-192, 1995.
- [3] W.W. Gaver, Technology Affordances, In *Proc. of CHI'91*, pp.79-84, ACM, 1991.
- [4] J.J. Gibson, *The Ecological Approach to Visual Perception*, Houghton Mifflin, New York, 1979.
- [5] S. Iga and M. Yasumura, A New Concept for Separating Real World and Virtual World, In *Proc. of WISS'95*, pp.58-66, 1995.
- [6] S. Iga and M. Yasumura, Real Object Remote Controller: Applying Real World Affordance to Computer Interfaces, In *Proc. of ICMI'96*, pp.18-22, 1996.
- [7] D.A. Norman, *The Psychology of Everyday Things*, Basic Books Inc., New York, 1988.
- [8] 佐々木正人, アフォーダンス — 新しい認知の理論, 岩波科学ライブラリー 12, 岩波書店, 1995.