

住宅VRシステムを用いた音声認識による映像操作機能の開発

—機能の概要と操作幅に関する評定実験—

小野 浩史*¹・本田 微笑*²・吉岡 優*³・青島 智恵*⁴・橋口 裕文*⁴・平手 小太郎*⁵

Keywords : VR, easy operation, voice recognition, change of material, change of daylight, evaluation

VR, 簡易操作, 音声認識, 素材変更, 昼夜変更, 評定実験

1. はじめに

筆者らは、仮想住環境シミュレーションシステムの開発を行い^{1)~2)}、住宅購入予定者に対して、本システムを活用したプレゼンテーションを実施してきた。プレゼンテーションを実施するに当たり、VRシステムの特徴でもある、体感者自身の操作に関して、検討をかさねた結果、音声認識技術を用いた簡易操作機能の開発に着手した。本報では、ロボット制御等で急速に進展している音声認識技術に着目し、音声によりコンピュータにアクションを起こすことで、VR表示・操作を実現する機能について開発の内容と操作性向上のために実施した評定実験についての概要を報告するものである。評定実験は音声認識によってVR映像がどれだけ平行移動・回転すればよいかを認識することを目的とし、最終的に操作幅を決定するためのシステム性能評価として位置づけた。

2. 音声認識による映像操作機能の開発

VR操作には様々なマンマシンインターフェース(以降:MMI)が提案され、現在最も汎用的なものとしてゲーム用JOYPADを使用したものが挙げられる。本システムにおいてもこのJOYPADを活用している。しかし、JOYPADの使用に当たっては、ある程度の習熟が必要であり、VR操作時の究極のMMIとは言えない。さらに筆者らは、空間認識に関する被験者実験を実施³⁾しており、その実験の結果のひとつとして、JOYPADを操作することは、体感者がVRに集中することを阻害する可能性が示唆された。そこで、より自然な操作に近づけるため音声認識を使用したVRシステムのプロトタイプを開発した。

* 1 技術センター建築技術開発部ニューフロンティア技術開発室
* 2 文化女子大学造形学部
* 3 昭和女子大学生活科学部
* 4 住宅事業本部商品開発部
* 5 東京大学大学院工学系研究科

2.1 音声認識ミドルウェア

音声認識と機器操作システム(ここではVRシステム)とを仲介するミドルウェアとして、UNMAGIC^{注1)}を採用した。ミドルウェアとして、オペレーションシステムとソフトウェアの中間に位置することで、ソフトウェアの立ち上げ等も実施できる。

2.2 VRシステムとの融合

図-1に音声認識の概要をイメージで示す。システムのプロトタイプとして、パソコン1台で音声認識とVR表示を行うタイプ(以降:共存タイプ)と、2台使用するタイプ(以降:独立タイプ)を製作した^{注2)}。図-2に独立タイプの音声認識の処理のデータフローを示す。なお新しく開発した機能は制御信号の受信部分とその信号を認識してVR表示・制御を行う部分である。共存タイプと独立タイプが考えられる。独立タイプの場合は、TCP/IPを使用した制御信号の送受信を行う。

2.3 音声認識による操作

表-1に音声認識による操作対象を示す。今回のプロトタイプ開発では、主な操作対象は5種類(ウォークスルー、平行移動、回転、各部屋移動、各部屋パターン変更)^{注3)}とした。特に、平行移動・回転に関してはJOYPADあるいはマウスによる微量変化に追従した音声操作による移動を再現することは難しく^{注4)}、その操作幅を決定する

注1 音声操作ミドルウェア「UNMAGIC」は、ボイスマジック社製である。このミドルウェアは、耐雑音性能として、平均的な事務所内の騒音レベルとされる60dB前後の環境下で認識率を維持しており、音声操作をするために予め自分の声を登録する必要がないという特徴を持っている。UNMAGICが採用するScanSoft(R)社の音声認識エンジンには「成人女性」、「成人男性」、「子供」の3種類のモデルが搭載されており、年齢や性別の制限されない音声操作が可能である。また、特定の操作をする操作語彙毎に登録をすることも可能である。

注2 共存タイプの場合VR表示のCPUに対する負荷が大きく、ミドルウェアである音声認識への影響もあり、認識率が低下する傾向にあった。また、共存タイプで、音声認識とVR表示をCPU毎に割り振る等の試みも行ったが、操作が煩雑になることから現実的ではなく最終的には独立タイプを採用した。



図-1 音声認識機能の概要 (イメージ)

Outline of the function for voice recognition (image)

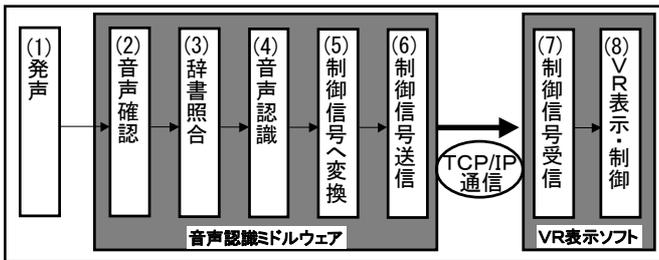


図-2 音声認識時の処理のデータフロー (独立タイプ)

Dataflow of processing for voice recognition (independent type)

表-1 音声認識による操作対象^{注3)}

Target of operation by voice recognition

No	操作対象	発声内容	具体的操作
1	ウォークスルー (開始・停止)	スタート・ストップ	あらかじめ準備されたウォークスルーの軌跡に沿って自動モードで視点移動を行う。音声認識をきっかけとし、操作が開始される。
2	平行移動 (前・後)	まえ・うしろ	仮想空間内にカメラモデルを設置し、そのカメラの姿勢制御を行う。カメラのローカル座標にて前後・左右・上下を判断する。発声確認後、各操作に対応した操作幅の移動を行う。操作幅に関しては12人の被験者に体感してもらい、最適な値を決定した。躯体・家具を対象に衝突判定することで、壁の通り抜けなどの不具合を解消した。
	平行移動 (右・左)	ライト・レフト	
	平行移動 (上・下)	アップ・ダウン	
3	回転 (右・左)	みぎ・ひだり	平行移動と同様、カメラモデルを用いて、カメラの姿勢制御を行う。カメラのローカル座標にて視点位置を固定し、音声認識後、各操作に対応した操作幅、参照点位置の回転を行う。
	回転 (上・下)	うえ・した	
4	各部屋移動	リビング 他	カメラモデルを各部屋に移動させる。各部屋には、ランダムにアクセス可能とする。
5	パターン変更 (昼・夜)	ひる・よる	あらかじめ準備されたテキストの切り替えを行う。音声認識をきっかけとし操作が開始される。切り替えるテキストのデータ量により、反応が異なる。因みに昼夜変更よりも床材変更の方がデータ量は少ない。
	パターン変更 (床)	ゆか	



図-3 音声認識による操作状況

Situation of operation with voice recognition

注3 発声内容はできるだけ単純な言葉とし、各発声内容が類似していないものを採用した。(右・左) (上・下) に関しては、内容が重なるので使用率の低いと思われる平行移動を英語表現とした。各部屋移動に関してはリビング、キッチン、寝室、2階を対象とした。

注4 音声認識で微量変化を再現するためには「もっと」「少し」等の修飾詞を認識させ、VR制御に反映すれば良いが、今

表-2 視点移動操作の操作幅に選定理由

Selected reasons of the operation width in the viewpoint transfer

操作対象	操作幅	操作幅の選定理由
平行移動 (前方)	20cm	住宅内はメーターモジュールが多くあり、一定の操作しか対応できない場合には不具合が生じると考え40cmを採用値とした。
	40cm	
	60cm	
平行移動 (後方)	10cm	前方の平行移動に比較して、後方は移動方向が判断できないため半分の操作幅とした。
	20cm	
	30cm	
平行移動 (右・左)	10cm	前方への平行移動と同様であるが、移動方向の映像情報がでないことから10cm減らした値を中央値とした。
	30cm	
	50cm	
平行移動 (上・下)	5cm	上・下移動はVR体感時には身長差の確認という場合が多く、最小操作幅を10cmとすることにした。
	10cm	
	15cm	
回転 (右・左)	10度	VR表示時に視野角90度を確保しており、45度以上の回転は映像の半分以上が変更することとなるので、最大操作幅を40度とした。
	25度	
	40度	
回転 (上・下)	5度	上・下への首振りには、住宅の場合それ程必要性がないが、吹き抜けや周辺環境の確認時を想定し、最大操作幅を5度とした。
	10度	
	15度	

ために12人の被験者による評価を実施した。図-3に音声認識による操作状況を示す。集音マイクに向かい発声することで操作が行われる^{注5)}。

2.4 視点移動操作の較正

表-1に視点移動操作幅と選定理由をまとめた。操作幅を検討する必要がある項目は平行移動4種類と回転2種類である。各々筆者らが最適と考える操作幅を選定し、その操作幅を中央値とした時の大小3寸法の操作幅に対して評価実験を実施し、最終的な操作幅を決定することとした。今回のプロトタイプシステムでは表-2中で陰影で示す操作幅を採用した。

2.5 音声認識による操作状況と考察

ウォークスルーに関しては、操作のタイミングをJOYPADのボタンで得るか音声認識で得るかの違いであり、体感者がJOYPAD操作に意識を阻害されることがないという点で有効と思われる。

平行移動、回転に関しては、操作幅の検討を行ったが、適切な操作幅は、映像の状況にも依存するものと考えられる。また「僅かな」操作あるいは、「大幅な」操作への対応を行っていないプロトタイプの範疇では、JOYPADの操作

回のプロトタイプでは対応していない。

注5 事務所空間での操作の場合でも、音声認識の誤認識は殆どない。また、集音マイクから体感者までの距離が5m離れても音声認識可能であった。ただし、複数人の同時発声の場合、誤認識する傾向にあった。集音マイクは2種類を検討した。付表にその2種類の形式と性能を示す。指向性を考慮して、最終的にはAT9820Xを採用した。

付表 集音マイクの性能

Performance of the directional microphones

名称	パウンダリー・マイクロフォン AT9750	コンパクト・スタンド・マイクロホン AT9820X
販売元	株式会社オーディオテクニカ	
形式	エレクトレット・コンデンサー型	バックエレクトレット・コンデンサー型
指向特性	単一指向性/無指向性切換え型	単一指向性型
周波数特性	100~10,000Hz	
感度(OdB=1V/Pa,1kHz)	-44±3dB	-38±3.5dB
インピーダンス	2kΩ	2kΩ
S/N比(1kHz/Pa)	55dB以上	56dB以上



図-4 時間の変更例
Change examples of time



図-5 床材の変更例
Change examples of floor material

性の方が優れる。

各部屋移動に関しては、具体的に部屋名称を告げることで、移動できる点は直感的であり、音声認識を用いる最も有効な状況である。

パターン変更に関しては、図-4に時間の変更例、図-5に床材の変更例を示す。JOYPADのボタン操作に比較して、操作対象を直接発声することで、各部屋移動と同様に直感的なイメージを誘発できる。

3. 評価実験

3.1 実験構成

図-6に実験構成を示す。評価対象が発声時のVR操作幅の妥当性に関するものであるため、事前の説明と操作の練習が必要である。また、操作時に随時評価する形式とすることでより精度の高いデータを収集した。表-3に体感パターンと操作幅の組み合わせを示す。表-4に被験者と体感パターンの組み合わせを示す。被験者は男性6名、女性6名とし、各(1)~(6)の組み合わせとした。図-7に評価対象の平面図を、図-8に各視点でのVR映像を示す。

3.2 評価方法

平行移動と回転に関して、操作幅を決定するために12人の被験者に3パターンの操作幅を組み合わせさせたコンテンツを各5分間体感させ、各体感後、各操作対象毎に満足度を10点満点で申告させた(ME法)。点数には基準が無いので、10点以上の申告も許容した。各パターンで同じ点数の場合は重ねて集計することとしたので、操作対象毎に必ずしも合計が12とはならない。ここでパターンA、B、Cの順序効果を考慮し、6名の被験者が異なる順序で体感

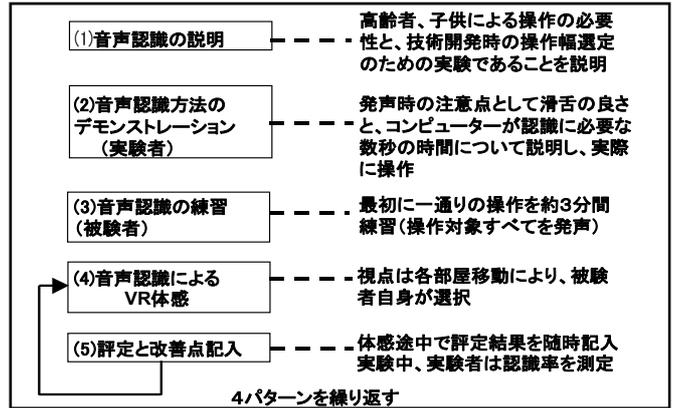


図-6 実験手順とポイント
Experiment process and some points

表-3 体感パターンと操作幅の組み合わせ

Combination of effective pattern and operation width

体感パターン	平行移動(前方)	平行移動(後方)	平行移動(右・左)	平行移動(上・下)	回転(右・左)	回転(上・下)
A	20cm	20cm	30cm	15cm	10度	15度
B	40cm	10cm	50cm	5cm	40度	10度
C	60cm	30cm	10cm	10cm	25度	5度
D	40cm	20cm	30cm	10cm	25度	10度

表-4 被験者と体感パターンの組み合わせ

Combination of subjects and effective pattern

被験者	(1)	(2)	(3)	(4)	(5)	(6)
体感順序	A→B→C	A→C→B	B→C→A	B→A→C	C→A→B	C→B→A

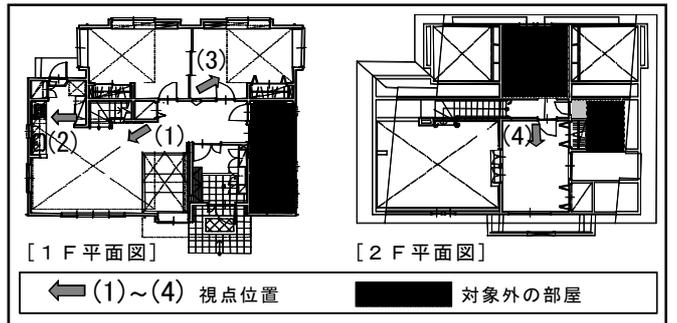


図-7 評価対象(平面)
A target space of evaluation (floor plans)



図-8 各視点位置でのVR映像
VR images in each viewpoint position

を行った。体感パターンDとは各操作幅の中央値であり、筆者らが最も有効と考えた組み合わせである。但し、体感

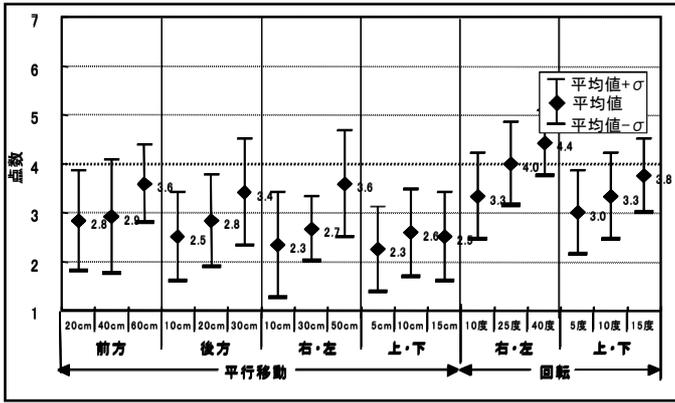


図-9 操作幅に関する評価 (全体)
Evaluation on the operation width (All subjects)

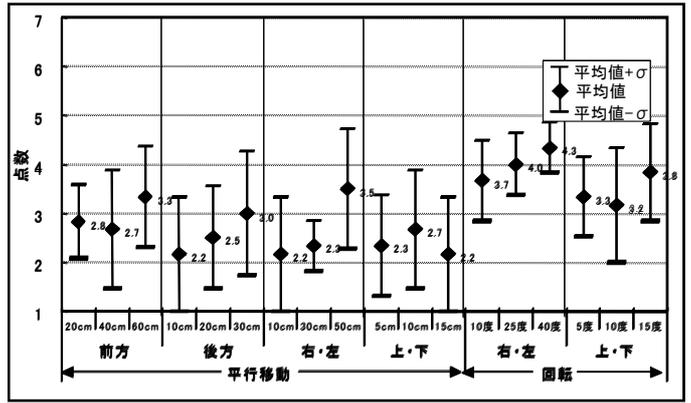


図-10 操作幅に関する評価 (男性)
Evaluation on the operation width (Men)

表-5 操作幅に関する多重比較結果

Results of Multiple Comparison on the operation width
多重比較:Fisherの最小有意差法 **: 1% 有意 *: 5% 有意

因子	水準 1	水準 2	平均値 1	平均値 2	差	P 値	判定
平行移動 (前方)	20	40	2.83	2.92	0.08	0.81	
	20	60	2.83	3.58	0.75	0.03	*
	40	60	2.92	3.58	0.67	0.06	
平行移動 (後方)	10	20	2.50	2.83	0.33	0.34	
	10	30	2.50	3.42	0.92	0.01	**
	20	30	2.83	3.42	0.58	0.10	
平行移動 (右・左)	10	30	2.33	2.67	0.33	0.34	
	10	50	2.33	3.58	1.25	0.00	**
	30	50	2.67	3.58	0.92	0.01	**
平行移動 (上・下)	5	10	2.25	2.58	0.33	0.34	
	5	15	2.25	2.50	0.25	0.47	
	10	15	2.58	2.50	0.08	0.81	
平行移動 (右・左)	10	25	3.33	4.00	0.67	0.06	
	10	40	3.33	4.42	1.08	0.00	**
	25	40	4.00	4.42	0.42	0.23	
平行移動 (上・下)	5	10	3.00	3.33	0.33	0.34	
	5	15	3.00	3.75	0.75	0.03	*
	10	15	3.33	3.75	0.42	0.23	

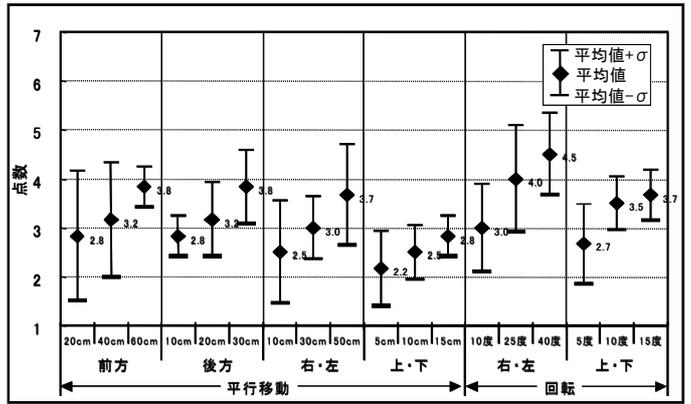


図-11 操作幅に関する評価 (女性)
Evaluation on the operation width (Women)

パターンDは集計には使用しておらず、参考データとした。更に、全ての体感後操作対象毎に7段階のSD法にて評価させ、改善点等の自由記述も実施した。SD法評価は操作幅が非常に少ない(1)から非常に多い(7)までとした。また、発声に伴う音声認識率の違いを実験者がカウントした。

3.3 操作幅に関する評価結果

図-9に操作幅に関する評価結果に関して、12人全員をまとめてチャートとして表示する。これらの結果はSD法による平均値を使用している。点数4が最適操作幅を意味する。筆者らが最適と考えた操作幅はすべて中央値であるが、その中央値がこれらの評価結果の中で最適と評価されたものは平行移動(上・下)と回転(右・左)だけであった。また操作幅が多いと申告したのは回転(右・左)の40度だけであった。各操作対象毎に操作幅間の多重比較を行った結果を表-5に示す。操作幅の最も小さいものと大きいものとの間に有意差があらわれる傾向であるが、平行移動(上・下)に関してはその傾向はみられない。また、平行移動(右・左)に関して、移動幅50cmが他の移動幅

よりも評価が高いこともわかる。

3.4 操作幅に関する評価結果の男女差

図-9を男性・女性と分けて、図-10、図-11に示す。全体的に男性の評価が女性の評価と比較して低いことがわかる。男性の評価と女性の評価が異なる操作対象は平行移動(上・下)、回転(上・下)である。

3.5 操作満足度に関する評価結果

図-12に全体を対象とした操作満足度に関する評価結果を示す。男性・女性の違いはなかった。満足度は操作幅のみでなく音声認識率の影響も加味された指標と考えられる。評価の結果、回転に関する満足度に差がないという結果が得られた。操作幅に関する評価と同様、平行移動(前方)(後方)(右・左)は最も大きな操作幅が最適と申告されている。

3.6 考察

図-13に発声カウントと認識率との関係を示す。各操作幅毎に全体・男性・女性と比較した。図の上部に認識率を示す。男性よりも女性の認識率が高い結果となった。また発声カウントでは回転の発生カウントが多く、特に女性が回転操作を多用する傾向にある。さらに前方への平行移動操作も女性に多用されている。傾向として女性の認識率が高かったが、個人的な滑舌の良し悪しにも影響されている

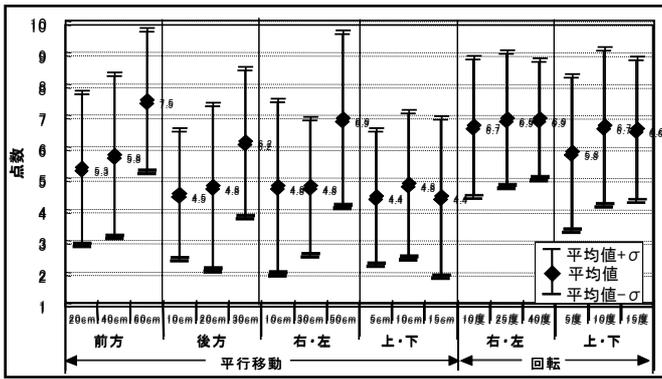


図-12 操作満足度に関する評定 (全体)
Evaluation on the operation satisfaction (All subjects)

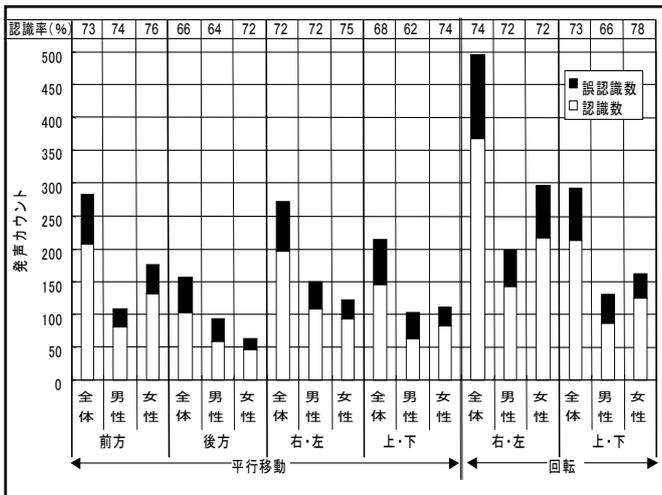


図-13 発声カウントと認識率の関係
Relationship between voice count and recognition rate

ものと思われた。

表-6に操作満足度を目的変数とした重回帰分析^{注6)}結果を示す。ここでは、各操作に関して各被験者毎に3回の操作幅評定結果の平均値と、誤認識率の平均値を求め、説明変数とした。説明変数は12項目に及ぶが、マルチコの問題により最終的には3項目を対象とした。決定係数は0.60と高く、分散分析の結果も5%の有意差が得られた。従って、重回帰分析結果の各係数により操作満足度の推定が可能となる。ここでは推定に意味があるのではなく、各

注6 操作満足度は各被験者毎の全操作に対する平均値を用い、操作幅評定結果は申告値4を100とし、申告値が1ずれる度にを75、50、25として集計した。

表-6 操作満足度に関する重回帰分析結果

Results of Multiple Regression Analysis on the operation satisfaction

回帰統計	
重相関 R	0.77
重決定 R2	0.60
補正 R2	0.45
標準誤差	1.24
観測数	12.00

決定係数が0.5より大きい

分散分析表		**:1%有意		*:5%有意	
	自由度	変動	分散	分散比	有意 F
回帰	3.00	18.54	6.18	3.99	0.05
残差	8.00	12.38	1.55		*
合計	11.00	30.93			

重回帰分析結果					
	係数	標準誤差	t	P-値	単相関係数
切片	3.99	2.18	1.83	0.10	—
回転(上・下)	0.05	0.02	2.20	0.06	0.55
回転(右・左)誤認識率	-0.03	0.03	-0.88	0.40	-0.43
回転(上・下)誤認識率	-0.05	0.02	-2.17	0.06	-0.44

説明変数の目的変数への寄与の割合に意味がある。重回帰分析結果の右欄に単相関係数を示す。誤認識率は逆相関となっており、回転(上・下)が操作満足度との関係において、最も単相関係数が高く、回転(上・下)の改善は、音声認識を用いた本機能の今後の課題と捉えられる。

4. おわりに

仮想住環境シミュレーションシステムの音声認識技術を用いた簡易操作機能の追加に関して、開発内容と操作性向上のために実施した検討項目に関して概要を述べた。また最適な操作幅を判断する目的で評定実験を実施し、評定の結果はシステム開発に反映された。今後はJOYPADと音声認識を共存し、より操作性の高いVRシステムへ拡張を行う予定である。

参考文献

- 1) 小野浩史他：仮想住環境シミュレーションシステムの開発，日本建築学会技術報告書，第14号，pp.171-176，2001.
- 2) 小野浩史他：集合住宅仮想体感システムの開発，日本建築学会技術報告集，第18号，pp.393-398，2003.
- 3) 小野浩史他：VRを用いた空間認識に関する基礎的研究その1～その5，日本建築学会大会（北海道）学術講演梗概集，2004.