

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第4709598号
(P4709598)

(45) 発行日 平成23年6月22日(2011.6.22)

(24) 登録日 平成23年3月25日(2011.3.25)

(51) Int.Cl.		F I	
G06T 13/40	(2011.01)	G06T 15/70	B
H04M 11/00	(2006.01)	H04M 11/00	302
H04M 1/00	(2006.01)	H04M 1/00	R

請求項の数 1 (全 34 頁)

(21) 出願番号	特願2005-201740 (P2005-201740)	(73) 特許権者	307020545
(22) 出願日	平成17年7月11日 (2005.7.11)		公立大学法人岡山県立大学
(65) 公開番号	特開2007-18433 (P2007-18433A)		岡山県総社市窪木111番地
(43) 公開日	平成19年1月25日 (2007.1.25)	(74) 代理人	100114535
審査請求日	平成20年6月13日 (2008.6.13)		弁理士 森 寿夫
		(74) 代理人	100075960
			弁理士 森 廣三郎
		(74) 代理人	100126697
			弁理士 松浦 瑞枝
		(72) 発明者	渡辺 富夫
			岡山県総社市窪木111番地 岡山県立大 学内
		(72) 発明者	山本 倫也
			岡山県総社市窪木111番地 岡山県立大 学内

最終頁に続く

(54) 【発明の名称】 アバター表示機能付き通話端末

(57) 【特許請求の範囲】

【請求項1】

身体リズムを本人に感得させるリズム画像を画面に表示する画像生成手段を備えたアバター表示機能付き通話端末であって、

画像生成手段は、音声取込部、動画取込部、信号取込部、アバター動画生成部、動画合成部、そして画像表示部から構成され、話し手動作及び聞き手動作を混在させて動く本人アバターの動画と相手の実写動画とを重ね合わせたリズム画像を生成してなり、

音声取込部は、本人及び相手の音声信号をアバター動画生成部に送り込み、

動画取込部は、相手の実写動画を動画合成部に送り込み、

信号取込部は、本人の操作信号をアバター動画生成部に送り込み、

アバター動画生成部は、本人の音声信号をON/OFF信号とみなし、呼気段落区分の各呼気段落に含まれるON信号の割合の移動平均として算出される本人アバターの話し手マクロ推定値が話し手マクロ閾値を超え、かつ本人の音声信号のサンプリング信号の移動平均として算出される本人アバターの話し手ミクロ推定値が話し手ミクロ閾値を超えた時点を本人アバターの話し手タイミングとして推定し、相手の音声信号をON/OFF信号とみなし、呼気段落区分の各呼気段落に含まれるON信号の割合の移動平均として算出される本人アバターの聞き手マクロ推定値が聞き手マクロ閾値を超え、かつ相手の音声信号のサンプリング信号の移動平均として算出される本人アバターの聞き手ミクロ推定値が聞き手ミクロ閾値を超えた時点を本人アバターの聞き手タイミングとして推定し、本人の操作信号が送り込まれた時点を本人アバターの特別タイミングとして、前記本人アバターの話し手タイミング、

10

20

聞き手タイミング及び特別タイミングを本人アバターのリズムタイミングとし、本人アバターのリズムタイミングが本人アバターの話し手タイミングであれば本人アバターの話し手動作を、本人アバターのリズムタイミングが本人アバターの聞き手タイミングであれば本人アバターの聞き手動作を、本人アバターのリズムタイミングが本人アバターの特別タイミングであれば特別動作を、それぞれ本人アバターのリズム動作として作り出し、前記本人アバターのリズムタイミングで本人アバターのリズム動作に従って動く本人アバターを後ろ向きにした動画を生成して、該本人アバターの動画を動画合成部に送り込み、動画合成部は、後ろ向きにした本人アバターの動画と前向きにした相手の実写動画とを対面関係に重ね合わせたリズム画像を生成して、該リズム画像を画像表示部に送り込み、画像表示部は、リズム画像を画面に表示することを特徴とするアバター表示機能付き通話 10
端末。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、隔地者として会話する本人及び相手の意思疎通を良好にするアバター表示機能付き通話端末に関する。

【0002】

本発明の「通話端末」は、無線又は有線等の通信手段により、隔地者である本人及び相手が互いに音声信号をやり取りする際に本人又は相手が用いる通信装置を意味し、固定電話、携帯電話や無線装置を含む。 20

【背景技術】

【0003】

会話する本人及び相手は、音声信号によるバーバル情報と音声信号以外のノンバーバル情報とを取得することにより、良好な意思疎通を図る。これから、例えば本人及び相手の音声信号のみを送受信する通常の携帯電話より、視覚的なノンバーバル情報として本人及び相手の実写動画を送受信できるテレビ電話機能付き携帯電話の方が、意思疎通に適していると考えられる。しかし、携帯電話は、詳細かつ円滑な本人及び相手の実写動画を送受信することが難しく、画面も小さいため、本人又は相手の実写動画からノンバーバル情報を取得しにくい。そこで、本人又は相手の代わりとして、実写動画よりデータ量又は計算量が少なく、円滑な動きができるアバター（代わりキャラクタ）を用いる携帯電話が提案さ 30
れている。

【0004】

例えば特許文献1は、本人又は相手に代えて、本人又は相手の音声信号に応答して動くアバター（CGキャラクタ）を用い、このアバターの動きから、本人又は相手にノンバーバル情報を取得させる携帯電話を提案している。ここで、前記アバターは、本人及び相手の音声信号から音声信号を解析し、本人又は相手の感情状態を推定して動きを制御しており、このアバターの動きを明確な感情表現として画面に表示させる。

【0005】

また、特許文献2は、音声信号の強弱に応じて動くアバター（キャラクタ）をアニメーション表示する携帯電話を提案している。更に、この特許文献2は、音声信号の強弱に応じてアバターが動くだけでは十分な感情表現ができないとして、携帯電話のキーを信号入力部（挿入部）として前記キーから操作信号を入力させることにより、本人又は相手の代わりとなるアバターに特定動作をさせている。前記特定動作は、本人又は相手が意図してアバターに実行させることから、アバターにより積極的な感情表現ができ、より明確なノンバーバル情報を本人又は相手に取得させることができると考えられる。 40

【0006】

【特許文献1】特開2003-248841号公報

【特許文献2】特開2005-064939号公報

【発明の開示】

【発明が解決しようとする課題】 50

【0007】

会話する本人及び相手の意思疎通は、音声信号によるバーバル情報が中心であるが、良好な意思疎通を実現するには、音声信号以外のノンバーバル情報が欠かせない。特許文献1及び特許文献2は、前記ノンバーバル情報の提供手段として動くアバターを用いている点で共通している。ところが、特許文献1の携帯電話は、アバターの動きがノンバーバル情報となるように本人又は相手の感情状態を推定する計算処理が必要で、実写動画よりデータ量又は計算量が少ないアバターを用いる利点を害している。また、本人及び相手の音声信号から人間の感情を推定することは難しく、果たして適切なノンバーバル情報をアバターの動きにより表現できるか疑問である。

【0008】

この点、特許文献2の携帯電話は、本人又は相手が意図的にアバターを操作し、自身の感情を直接アバターの動きとして表現できるため、より適切なノンバーバル情報が提供できると考えられる。しかし、本人又は相手が操作しない間、アバターは単に音声信号の強弱にตอบสนองして動くのみで、時折本人又は相手がアバターを操作するだけでは、アバターの動きによるノンバーバル情報の取得が間断的になりすぎて、適切なノンバーバル情報をアバターの動きにより表現できるとは考えにくい。

【0009】

相手と会話する本人は、会話の流れにおける話す又は聞くといったタイミングを節としたリズムを感得し、前記リズムからノンバーバル情報を取得することにより、良好な意思疎通を実現する。ここで、前記話す又は聞くといったリズムが「会話における本人の身体リズム（以下、身体リズムと略）」である。上記特許文献1及び特許文献2は、こうした身体リズムに考慮が払われていない。しかし、本人又は相手に代わるアバターを用いて意思疎通を図る場合、アバターの動きにより身体リズムを感得させることが、本人にノンバーバル情報を取得させる最善の手段である。そこで、本人又は相手に代わるアバターにより会話における本人の身体リズムを本人に感得させることにより、本人にノンバーバル情報を取得させる携帯電話、固定電話や無線装置等の通話端末を開発するため、検討した。

【課題を解決するための手段】

【0010】

検討の結果開発したものが、隔地者である本人及び相手が会話する際に本人が用いる通話端末であって、画面に本人アバターを表示するアバター表示機能を付加した通話端末において、本人及び相手の音声信号から推定される話し手タイミング及び聞き手タイミングや本人の操作信号から特定される特別タイミングを本人アバターのリズムタイミングとし、このリズムタイミングで実行するリズム動作として本人アバターの話し手動作及び聞き手動作を作り出し、前記リズムタイミングでリズム動作に従って動く本人アバターの動画を相手の実写動画と重ね合わせたりズム画像を生成して、このリズム画像を画面に表示する画像生成手段を備えてなり、画面に表示されるリズム画像より会話における本人の身体リズムをこの本人に感得させることにより、この本人にノンバーバル情報を取得させるアバター表示機能付き通話端末である。

【0011】

本発明の通話端末は、本人及び相手の音声信号から推定される話し手タイミング及び聞き手タイミングを本人アバターのリズムタイミングとし、話し手を表現する話し手動作と聞き手を表現する聞き手動作とから構成されるリズム動作を、前記リズムタイミングで実行することにより、本人アバターを動かす。すなわち、本発明の本人アバターは、本人又は相手の代わりとなるばかりでなく、本人に身体リズムを感得させる媒体として働く。こうして、本人アバターの動画を用いたリズム画像を見る本人は、このリズム画像から身体リズムを感得し、このリズム画像からノンバーバル情報を取得できる。ここで、本発明の「音声信号」は、本人又は相手が発する生の音声を携帯電話に取り込んで処理される連続した電気信号（アナログ信号）を意味する。

【0012】

「本人アバター」は、本人の代わりとなって動くキャラクタを意味する。この本人アバタ

10

20

30

40

50

一は、人間を簡略化又は抽象化（デフォルメ）したキャラクタを基本とするが、擬人化されていれば、動物や植物のほか、自動車や飛行機の乗り物、建物、樹木、草花、岩や、更に空想上の生物又は無生物でもよい。これら本人アバターは、人間の頭、口、目、腕、胴又は脚や動物の耳、尻尾に相当する可動部を有し、各可動部を直線運動又は円運動動させて、動きを作り出す。

【0013】

「話し手タイミング」は、話し手として本人アバターが動くタイミングを、また「聞き手タイミング」は、聞き手として本人アバターが動くタイミングを意味し、特に聞き手タイミングは聞き手として本人アバターが頭の頷き動作をするタイミングに相当する。「リズムタイミング」は、身体リズムを表現するために本人アバターが動くタイミングを意味し、前記話し手タイミング及び聞き手タイミングに、後述する特別タイミングを加えている。

10

【0014】

「話し手動作」は、話し手として動く本人アバターの動きを、また「聞き手動作」は、聞き手として動く本人アバターの動きを意味する。「リズム動作」は、身体リズムを表現するために動く本人アバターの動きを意味し、前記話し手動作及び聞き手動作のほか、後述する特別動作を加える。「特別動作」は、本人又は相手の意図又は意思表示を伝達する本人アバターの特定の動きであり、特にリズム動作に対して有用な強い肯定反応を表す動作、すなわちゆっくり又は早くて深い頭の頷き動作や、強い否定反応を表す動作、すなわちゆっくり又は早くて大きな頭の（左右）振り動作が好ましいが、例えばピースサイン、万歳や別れを示す手を振る動作でもよい。「本人アバターの動画」は、リズム動作に従って動く本人アバターの動画を意味する。

20

【0015】

「リズム画像」は、上記本人アバターの動画を含んだ表示用の画像で、上述したように、本人に身体リズムを感得させ、更にノンバーバル情報を取得させる媒体を意味する。ここで、本発明の本人アバターは、それぞれが身体リズムを表す話し手動作又は聞き手動作に従って動くため、本人アバターの動画のみでも、本人に身体リズムを感得させる働きを有する。しかし、会話は本人及び相手が共に存在して成立するものであるから、これら当事者の一方の代わりとなる本人アバターの動画を単に表示するだけでは、身体リズムを本人に感得させにくい場合も考えられる。例えば、通話端末が携帯電話である場合、本人アバターの動画は相当小さくなり、本人アバターが当事者の一方の代わりとして動いていると認識しづらくなる。

30

【0016】

そこで、本発明は、話し手動作及び聞き手動作を混在させて動く本人アバターの動画と相手の実写動画とを重ね合わせたリズム画像を用いて、リズム画像を見る本人がノンバーバル情報を取得しやすくした。このほか、話し手動作、聞き手動作及び特別動作を混在させて動く相手アバターの動画からなるリズム画像や、話し手動作及び聞き手動作を混在させて動く本人アバターの動画と話し手動作及び聞き手動作を混在させて動く相手アバターの動画とを重ね合わせたリズム画像を用いることも考えられる。

【0017】

話し手動作及び聞き手動作を混在させて動く本人アバターの動画と相手の実写動画とを重ね合わせたリズム画像を用いる通話端末は、画像生成手段は、音声取込部、動画取込部、信号取込部、アバター動画生成部、動画合成部、そして画像表示部から画像生成手段を構成し、リズム画像を生成する。音声取込部は、本人及び相手の音声信号をアバター動画生成部に送り込む。動画取込部は、相手の実写動画を動画合成部に送り込む。信号取込部は、本人の操作信号をアバター動画生成部に送り込む。アバター動画生成部は、本人の音声信号から本人アバターの話し手タイミングを推定し、また相手の音声信号から本人アバターの聞き手タイミングを推定し、本人の操作信号が送り込まれた時点をも本人アバターの特別タイミングとして、前記本人アバターの話し手タイミング、聞き手タイミング及び特別タイミングを本人アバターのリズムタイミングとし、本人アバターのリズムタイミングが

40

50

本人アバターの話し手タイミングであれば本人アバターの話し手動作を、本人アバターのリズムタイミングが本人アバターの聞き手タイミングであれば本人アバターの聞き手動作を、本人アバターのリズムタイミングが本人アバターの特別タイミングであれば特別動作を、それぞれ本人アバターのリズム動作として作り出し、前記本人アバターのリズムタイミングで本人アバターのリズム動作に従って動く本人アバターを後ろ向きにした動画を生成して、この本人アバターの動画を動画合成部に送り込む。動画合成部は、後ろ向きにした本人アバターの動画と前向きにした相手の実写動画とを対面関係に重ね合わせたリズム画像を生成して、このリズム画像を画像表示部に送り込む。そして、画像表示部は、リズム画像を画面に表示する。前記アバター動画生成部が生成する本人アバターの動画は、話し手動作及び聞き手動作が不規則に混在し、場合によって重複して複雑な動きを形成するため、全体としてより自然な本人アバターのリズム動作を作り出すことができる。

10

【0018】

本人の音声信号は、本人が用いる通話端末の内蔵マイク又は通話端末に接続した外部マイクから音声取込部に取り込まれる。相手の音声信号は、相手が用いる通話端末の内蔵マイク又は通話端末に接続した外部マイクから通話端末の送受信部により発信され、本人が用いる通話端末の送受信部に受信されて、音声取込部に取り込まれる。相手の実写動画は、相手が用いる通話端末の内蔵カメラ又は通話端末に接続した外部カメラから通話端末の送受信部により発信され、本人が用いる通話端末の送受信部に受信されて、動画取込部に取り込まれる。動画合成部は、本人アバターの動画と相手の実写動画とを合成して1枚のリズム画像を生成する。ここで、リズム画像における本人アバターと相手とは対面する位置関係、例えば本人アバターは顔が見える程度の斜視に後ろ向きで画面右下に配し、相手は前向きで画面中央に配すると、対面者として本人アバター及び相手が会話しているように見え、このリズム画像を見る本人に身体リズムを感得させやすくなる。

20

【0019】

「本人の音声信号から本人アバターの話し手タイミングを推定」するとは、本人の音声信号の変化から本人アバターの話し手タイミングの存在の有無を判断し、前記存在が肯定された時点を話し手タイミングとすることを意味する。例えば本人の音声信号の移動平均を算出し、前記算出された結果が推定閾値を超えた場合に本人アバターの話し手タイミングの存在を肯定し、この肯定判断がされた時点を本人アバターの話し手タイミングとする。ここで、前記移動平均には、線形結合、非線形結合又はニューラルネットワーク結合等を利用しうる。「相手の音声信号から本人アバターの聞き手タイミングを推定」する意味も同様である。

30

【0020】

本発明の応用として、話し手動作、聞き手動作及び特別動作を混在させて動く相手アバターの動画からなるリズム画像を用いる通話端末は、画像生成手段を音声取込部、信号取込部、アバター動画生成部、そして画像表示部から構成し、前記リズム画像を生成する。音声取込部は、本人及び相手の音声信号をアバター動画生成部に送り込む。信号取込部は、相手の操作信号をアバター動画生成部に送り込む。アバター動画生成部は、相手の音声信号から相手アバターの話し手タイミングを推定し、本人の音声信号から相手アバターの聞き手タイミングを推定し、また相手の操作信号が送り込まれた時点を相手アバターの特別タイミングとして、前記相手アバターの話し手タイミング、聞き手タイミング及び特別タイミングを相手アバターのリズムタイミングとし、相手アバターのリズムタイミングが相手アバターの話し手タイミングであれば相手アバターの話し手動作を、相手アバターのリズムタイミングが相手アバターの聞き手タイミングであれば相手アバターの聞き手動作を、相手アバターのリズムタイミングが相手アバターの特別タイミングであれば相手アバターの特別動作を、それぞれ相手アバターのリズム動作として作り出し、前記相手アバターのリズムタイミングで相手アバターのリズム動作に従って動く相手アバターの動画を生成して、この相手アバターの動画を画像表示部に送り込む。そして、画像表示部は、リズム画像を画面に表示する。前記アバター動画生成部が生成する相手アバターの動画は、話し手動作、聞き手動作及び特別動作が不規則に混在し、場合によって重複して複雑な動きを

40

50

形成するため、全体としてより自然な相手アバターのリズム動作を作り出すことができる。

【0021】

本人及び相手の音声信号の取り込みや、相手アバターの話し手タイミング及び聞き手タイミングの推定は、既述したところと同様であるため、説明を省略する。相手の操作信号は、相手が用いる通話端末の操作入力部又は通話端末に接続した外部入力部から通話端末の送受信部により発信され、本人が用いる通話端末の送受信部に受信されて、信号取込部に取り込まれる。ここで、操作入力部は通話端末が備えるテンキーを例示できる。また、外部入力部は、通話端末に接続してデータ入力に利用できるテンキーパッド、ジョイスティック、ゲームコントローラやデータグローブ等を例示できる。

10

【0022】

ここで、上記通話端末から通信機能を省略し、本人アバターの話し手動作、聞き手動作及び特別動作を混在させて動く本人アバターの動画からなるリズム画像を相手に見せれば、このリズム画像から身体リズムを本人及び相手に感得させ、本人及び相手にノンバーバル情報を取得させることができる。こうしたリズム画像を画面に表示する装置は、通話端末ではなく、「会話支援端末」と呼ぶことができる。

【0023】

本発明の応用である会話支援端末は、画像生成手段を音声取込部、信号取込部、アバター動画生成部、そして画像表示部から構成し、本人アバターの話し手動作、本人アバターの聞き手動作及び特別動作を混在させて動く本人アバターの動画からなるリズム画像を生成する。音声取込部は、本人及び相手の音声信号をアバター動画生成部に送り込む。信号取込部は、本人の操作信号をアバター動画生成部に送り込む。アバター動画生成部は、本人の音声信号から本人アバターの話し手タイミングを推定し、相手の音声信号から本人アバターの聞き手タイミングを推定し、また本人の操作信号が送り込まれた時点を本人アバターの特別タイミングとして、前記本人アバターの話し手タイミング、聞き手タイミング及び特別タイミングを本人アバターのリズムタイミングとし、本人アバターのリズムタイミングが本人アバターの話し手タイミングであれば本人アバターの話し手動作を、本人アバターのリズムタイミングが本人アバターの聞き手タイミングであれば本人アバターの聞き手動作を、本人アバターのリズムタイミングが本人アバターの特別タイミングであれば本人アバターの特別動作を、それぞれ本人アバターのリズム動作として作り出し、前記本人アバターのリズムタイミングで本人アバターのリズム動作に従って動く本人アバターの動画を生成して、この本人アバターの動画を画像表示部に送り込む。そして、画像表示部は、リズム画像を画面に表示する。

20

30

【0024】

本人及び相手の音声信号の取り込みや、本人アバターの話し手タイミング及び聞き手タイミングの推定は、既述したところと同様であるため、説明を省略する。本人の操作信号は、本人が用いる通話端末の操作入力部又は通話端末に接続した外部入力部から入力され、信号取込部に取り込まれる。具体的な操作入力部又は外部入力部は、既述したところと同様であるため、説明を省略する。

【0025】

また、本発明の応用として、話し手動作及び聞き手動作を混在させて動く本人アバターの動画と話し手動作及び聞き手動作を混在させて動く相手アバターの動画とを重ね合わせたリズム画像を用いる通話端末は、画像生成手段を音声取込部、アバター動画生成部、動画合成部、そして画像表示部から構成し、前記リズム画像を生成する。音声取込部は、本人及び相手の音声信号をアバター動画生成部に送り込む。アバター動画生成部は、本人の音声信号から本人アバターの話し手タイミングを推定し、また相手の音声信号から本人アバターの聞き手タイミングを推定して、前記本人アバターの話し手タイミング及び聞き手タイミングを本人アバターのリズムタイミングとし、本人アバターのリズムタイミングが本人アバターの話し手タイミングであれば本人アバターの話し手動作を、本人アバターのリズムタイミングが本人アバターの聞き手タイミングであれば本人アバターの聞き手動作を

40

50

、それぞれ本人アバターのリズム動作として作り出し、前記本人アバターのリズムタイミングで本人アバターのリズム動作に従って動く本人アバターの動画を生成して、この本人アバターの動画を動画合成部に送り込むと共に、相手の音声信号から相手アバターの話し手タイミングを推定し、また本人の音声信号から相手アバターの聞き手タイミングを推定して、前記相手アバターの話し手タイミング及び聞き手タイミングを相手アバターのリズムタイミングとし、相手アバターのリズムタイミングが相手アバターの話し手タイミングであれば相手アバターの話し手動作を、相手アバターのリズムタイミングが相手アバターの聞き手タイミングであれば相手アバターの聞き手動作を、それぞれ相手アバターのリズム動作として作り出し、前記相手アバターのリズムタイミングで相手アバターのリズム動作に従って動く相手アバターの動画を生成して、この相手アバターの動画を動画合成部に送り込む。動画合成部は、本人アバターの動画と相手アバターの動画とを重ね合わせたりリズム画像を生成して、このリズム画像を画像表示部に送り込む。そして、画像表示部は、リズム画像を画面に表示する。

10

【0026】

本人及び相手の音声信号の取り込みや、本人アバター及び相手アバターの話し手タイミング及び聞き手タイミングの推定は、既述したところと同様であるため、説明を省略する。動画合成部は、本人アバターの動画と相手アバターの動画とを合成して1枚のリズム画像を生成する。ここで、リズム画像における本人アバターと相手アバターとは対面する位置関係、例えば本人アバターは顔が見える程度の斜視に後ろ向きで画面右下に配し、相手アバターは前向きで画面中央に配すると、対面者として本人アバター及び相手アバターが会話しているように見え、本人に身体リズムを感得させやすくなる。

20

【0027】

リズム画像は、本人アバターの動画と相手の実写動画との組み合わせや、本人アバターと相手アバターとの組み合わせにより、本人に身体リズムを感得させ、ノンバーバル情報を取得させる。しかし、本人アバター又は相手アバターに特別動作をさせると、本人に身体リズムをより感得させやすく、また本人にノンバーバル情報を取得させやすくなる。

【0028】

特別動作は、画像生成手段に信号取込部を追加して生成する。信号取込部は、本人の操作信号をアバター動画生成部に送り込む。アバター動画生成部は、本人の音声信号から本人アバターの話し手タイミングを推定し、相手の音声信号から本人アバターの聞き手タイミングを推定し、また本人の操作信号が送り込まれた時点を本人アバターの特別タイミングとして、前記本人アバターの話し手タイミング、聞き手タイミング及び特別タイミングを本人アバターのリズムタイミングとし、本人アバターのリズムタイミングが本人アバターの話し手タイミングであれば本人アバターの話し手動作を、本人アバターのリズムタイミングが本人アバターの聞き手タイミングであれば本人アバターの聞き手動作を、本人アバターのリズムタイミングが本人アバターの特別タイミングであれば特別動作を、それぞれ本人アバターのリズム動作として作り出し、前記本人アバターのリズムタイミングで本人アバターのリズム動作に従って動く本人アバターの動画を生成して、この本人アバターの動画を画像合成部に送り込む。本人及び相手の音声信号の取り込みや、本人の操作信号の取り込みや、本人アバターの話し手タイミング及び聞き手タイミングの推定は、既述したところと同様であるため、説明を省略する。

30

40

【0029】

同様に、相手アバターの特別動作は、画像生成手段に信号取込部を追加して生成する。信号取込部は、相手の操作信号をアバター動画生成部に送り込む。アバター動画生成部は、相手の音声信号から相手アバターの話し手タイミングを推定し、本人の音声信号から相手アバターの聞き手タイミングを推定し、また相手の操作信号が送り込まれた時点を相手アバターの特別タイミングとして、前記相手アバターの話し手タイミング、聞き手タイミング及び特別タイミングを相手アバターのリズムタイミングとし、相手アバターのリズムタイミングが相手アバターの話し手タイミングであれば相手アバターの話し手動作を、相手アバターのリズムタイミングが相手アバターの聞き手タイミングであれば相手アバターの

50

聞き手動作を、相手アバターのリズムタイミングが相手アバターの特別タイミングであれば相手アバターの特別動作を、それぞれ相手アバターのリズム動作として作り出し、前記相手アバターのリズムタイミングで相手アバターのリズム動作に従って動く相手アバターの動画を生成して、この相手アバターの動画を画像表示部に送り込む。本人及び相手の音声信号の取り込みや、相手の操作信号の取り込みや、相手アバターの話し手タイミング及び聞き手タイミングの推定は、既述したところと同様であるため、説明を省略する。

【0030】

本発明の各通話端末は、身体リズムを表すリズムタイミングでリズム動作に従って動く本人アバターの動画を相手の実写動画と組み合わせたりリズム画像を本人に見せることにより、本人に身体リズムを感得させ、ノンバーバル情報を取得させる点に特徴がある。すなわち、本人アバターが動くリズムタイミングと、このリズムタイミングで実行するリズム動作とが重要である。

10

【0031】

まず、話し手タイミングは、次のように推定する。すなわち、アバター動画生成部は、本人の音声信号から算出される本人アバターの話し手マクロ推定値が本人話し手マクロ閾値を超えた場合のマクロ推定と、本人の音声信号から算出される本人アバターの話し手ミクロ推定値が本人話し手ミクロ閾値を超えた場合のミクロ推定とが共に肯定された時点を本人アバターの話し手タイミングとして推定する。ここで、話し手マクロ推定値は、本人の音声信号をON/OFF信号とみなし、呼気段落区分の各呼気段落に含まれるON信号の割合の移動平均として算出し、本人話し手ミクロ推定値は、本人の音声信号のサンプリング信号の

20

【0032】

同様に、聞き手タイミングは次のように推定する。すなわち、アバター動画生成部は、相手の音声信号から算出される本人アバターの聞き手マクロ推定値が本人聞き手マクロ閾値を超えた場合のマクロ推定と、相手の音声信号から算出される本人アバターの聞き手ミクロ推定値が本人聞き手ミクロ閾値を超えた場合のミクロ推定とが共に肯定された時点を本人アバターの聞き手タイミングとして推定する。ここで、聞き手マクロ推定値は、相手の

30

【0033】

「本人の音声信号（相手の音声信号）をON/OFF信号とみなすとは、アナログ信号である本人の音声信号（相手の音声信号）に閾値を設定し、前記閾値を境としてON及びOFFを繰り返すデジタル信号として、信号処理することを意味する。「呼気段落区分の各呼気段落に含まれるON信号の割合の移動平均」とは、人間が一息に発する音声信号の区分（＝呼気段落）である各呼気段落に含まれるON信号の割合を、前記段落単位で移動平均することを意味する。呼気段落区分の時間長は長短様々である。また、「サンプリング信号の移動平均」とは、音声信号を一定間隔でサンプリングしたサンプリング信号の強度（＝音声信号の大きさ）を、前記サンプリング単位で移動平均することを意味する。各移動平均は、既述したところと同様、線形結合、非線形結合又はニューラルネットワーク結合により算出する。

40

【0034】

ここで、上述の本人アバターの話し手タイミング及び聞き手タイミングの推定は、本人及び相手の音声信号のいずれから推定するかの違いだけで、上述から明らかなように、いずれも同じアルゴリズムを用いている。しかし、本人又は相手の音声信号の違いのほか、例えば線形結合による移動平均に用いる各係数を異ならせたり、存在の有無を判断する各閾

50

値を異ならせることにより、本人アバターの話し手タイミング及び聞き手タイミングは、それぞれ違ったタイミングとなる。また、例えば本人アバターの可動部を複数のグループに分け、各グループ毎に話し手タイミング及び聞き手タイミングを異ならせることもできる。ここで、前記線形結合における移動平均に用いる各係数や各閾値は、各リズム画像を見る本人の官能評価試験から、実証的に最適化していくとよい。

【0035】

本人アバターの話し手動作、聞き手動作及び特別動作は、次のように作り出すとよい。すなわち、アバター動画生成部は、本人アバターの可動部である口の開閉動作と、この本人アバターの前記口を除くその他の可動部のランダムな身体動作とを組み合わせ、本人アバターの話し手動作を作り出す。この場合、アバター動画生成部は、話し手動作における口の開閉動作の話し手タイミングと、話し手動作における前記口を除くその他の可動部のランダムな身体動作の話し手タイミングとを個別に導き出してもよい。ここで、「口を除くその他の可動部のランダムな身体動作」とは、本人アバターに設定された可動部のうち、口を除く可動部、例えば頭、目、腕、胴又は脚等の中から乱数に従って選択した単数又は複数の可動部を同時に動かして構成される本人アバター全体の動きを意味する。

10

【0036】

上記話し手動作は、口の開閉動作と、口を除くその他の可動部のランダムな身体動作とを同時に作り出し、同じ話し手タイミングで同時に実行させる場合を基本とする。これに対し、口の開閉動作の話し手タイミングと、口を除くその他の可動部の話し手タイミングとを個別に推定する場合、口の開閉動作と、口を除くその他の可動部のランダムな身体動作とは異なる話し手タイミングで個別に実行させる。これは、口を除くその他の可動部のランダムな身体動作の話し手タイミングを推定する判断基準となる各閾値を、口の開閉動作の話し手タイミングを推定する判断基準となる各閾値より低くすることで、口を除くその他の可動部のランダムな身体動作を口の開閉動作に比べて活発化させ、本人アバターを積極的に動くように見せることができる。

20

【0037】

また、アバター動画生成部は、本人アバターの可動部である頭の傾き動作と、この本人アバターの前記頭を除くその他の可動部をランダムに選択した身体動作とを組み合わせ、本人アバターの聞き手動作を作り出す。この場合、アバター動画生成部は、聞き手動作における頭の傾き動作の話し手タイミングと、聞き手動作における前記頭を除くその他の可動部をランダムに選択した身体動作の話し手タイミングとを個別に導き出してもよい。ここで、「頭を除くその他の可動部のランダムな身体動作」とは、本人アバターに設定された可動部のうち、頭を除く可動部、例えば目、口、腕、胴又は脚等の中から選択した単数又は複数可動部を同時に動かして構成される本人アバター全体の動きを意味する。この聞き手動作における頭の傾き動作と、頭を除くその他の可動部をランダムに選択した身体動作とを、同時に実行させる場合や、個別に実行させる場合は、上記話し手動作の例と同様であるため、説明は省略する。

30

【0038】

更に、アバター動画生成部は、本人アバターの可動部を予め選択した身体動作からなる特別動作を、本人アバターのリズム動作として作り出す。ここで、「可動部を予め選択した身体動作」とは、本人アバターに設定された可動部、例えば頭、目、口、腕、胴又は脚等のうち、予め選択した単数又は複数可動部を同時に動かして構成される本人アバター全体の動きを意味する。この特別動作は、本人又は相手が意図的に本人アバターを動かす場合の動きであるから、可動部を複数のグループに分けて個別に特別タイミングを導き出すことはないが、予め選択した可動部それぞれを、予め設定した時間差で動かすようにしてもよい。

40

【0039】

本人アバターは人間又は擬人化されたキャラクタであるから、可動部として頭、口、目、腕、胴又は脚等を用いることができる。前記可動部のうち、頭、腕、胴又は脚等の各動作は、関節を軸として、所定角度で各可動部が前後、左右又は上下に往復する揺動運動（往

50

復回動運動)として構成できる。具体的な動作として、頭の頷き動作は、首を軸として、所定角度で頭が前後に往復する揺動運動(往復回動運動)として構成できる。また、口の開閉動作は、本人又は相手の音声信号の強弱に応じた口の開閉運動として構成できる。そして、目の瞬き動作は、本人又は相手の音声信号と無関係に、ランダムな目の開閉運動として構成できる。

【発明の効果】

【0040】

本発明は、隔地者である本人及び相手が通話端末を用いた会話する際に、通話端末の画面に本人アバターを用いたリズム画像を表示することにより本人に身体リズムを感得させて、ノンバーバル情報を本人に取得させる効果を有する。ここで、本人に身体リズムを感得させる効果は、話し手動作及び聞き手動作を混在させて動く本人アバターの動画と相手の実写動画とを重ね合わせたリズム画像を本人に見せることにより得られる効果である。

10

【0041】

話し手動作及び聞き手動作を混在させて動く本人アバターの動画と相手の実写動画とを重ね合わせたリズム画像は、身体リズムに従って動く本人アバターが相手と同一空間内、すなわち画面内に表示されることにより、本人アバターと相手との疑似的な会話を感じさせて本人に身体リズムを感得させやすくし、この本人にノンバーバル情報を取得させる。これから、本人アバターと相手とは対面位置関係に配置されていると、前記疑似的な会話をより感得させやすく、本人に与えるノンバーバル情報も多くできる。

【0042】

本発明の応用として、話し手動作、聞き手動作及び特別動作を混在させて動く相手アバターの動画からなるリズム画像は、身体リズムに従って動く相手アバターが相手の意図的な操作を表す特別動作で動くことにより、本人に対する相手アバターの関わり合いが緊密であると思わせて本人に身体リズムを感得させ、この本人にノンバーバル情報を取得させる。ここで、相手アバターの特別動作は、相手が特に本人に伝達したい意図を身体リズムに組み入れる働きを有し、感得する身体リズムから本人に与えるノンバーバル情報を多くする。また、この(2)話し手動作、聞き手動作及び特別動作を混在させて動く相手アバターの動画からなるリズム画像を応用した会話支援端末は、特に健常者と会話する身体障害者の前記健常者に対する意思表示又は感情表現を補い、健常者と身体障害者との会話を円滑にする効果をもたらす。

20

30

【0043】

また、本発明の応用として、話し手動作及び聞き手動作を混在させて動く本人アバターの動画と話し手動作及び聞き手動作を混在させて動く相手アバターの動画とを重ね合わせたリズム画像は、上記リズム画像を組み合わせた効果を有する。すなわち、身体リズムに従って動く本人アバター及び相手アバターの疑似的な会話や、本人及び相手の身体リズムをそれぞれのリズム動作によって明示する本人アバター及び相手アバターの関わり合いが緊密であることを感得させて本人に身体リズムを感得させ、この本人にノンバーバル情報を取得させる。更に、本人アバター又は相手アバターに特別動作又は相手アバターの特別動作が加われば、本人に対する本人アバター又は相手アバターの関わり合いも緊密になり、本人に与えるノンバーバル情報も多くできる。

40

【発明を実施するための最良の形態】

【0044】

以下、本発明の実施形態について図を参照しながら説明する。図1は本人アバター51の動画と相手53の実写動画とを重ね合わせたリズム画像5を用いた携帯電話1の使用状態を示す斜視図であり、図2は本人アバター51の動画と相手53の実写動画とを重ね合わせたリズム画像5を用いた携帯電話1の本発明に係る構成を示すブロック図である(本人アバター51及び相手53は、後掲図13参照)。図2をはじめとするブロック図では、説明の便宜上、本発明に直接関係しない携帯電話の機能に関する部分は図示を省略し、また必要な部分も送受信部12により代表している(以下、同じ)。

【0045】

50

本発明を適用した携帯電話1の構成と、リズム画像5の生成手順について説明する。本例は、本人アバター51の動画と相手53の実写動画とを重ね合わせたリズム画像5を用いた携帯電話1の例である。これから、図1に見られるように、本人及び相手が用いる各携帯電話1は、互いに本人及び相手の音声信号を送受信するほか、互いに本人及び相手の実写動画を送受信する。ここで、本人の音声信号は相手から見れば相手の音声信号、本人の実写動画は相手から見れば相手の実写動画になる。本人及び相手が用いる携帯電話1は同じである。よって、以下では本人が用いる携帯電話1について説明し、各部の説明する場合は「本人が用いる携帯電話」の修飾句を省略して符号を付し、相手が用いる携帯電話1の各部について触れる場合は「相手が用いる携帯電話」と修飾句を付加して符号を省略する。

【0046】

携帯電話1は、図2に見られるように、音声取込部13、動画取込部14、信号取込部15、アバター動画生成部16、動画合成部17、そして画像表示部18から、本発明に係る主要部を構成する。この主要部は、従来より携帯電話に搭載されている情報処理装置をプログラムにより制御し、ソフト的に構成できる。ここで、例えば音声取込部13、動画取込部14、信号取込部15、動画合成部17及び画像表示部18は、処理内容が特定されているので、専用LSI等を用いてハード的に構成してもよい。図2中、本人に関係する信号は実線により、相手に関係する信号は波線により、そしてリズム画像の信号は太実線により、それぞれ図示している。

【0047】

音声取込部13は、内蔵マイク112から本人の音声信号を取り込む。前記本人の音声信号は、内蔵マイク112から送受信部12に取り込まれて発信され、相手が用いる携帯電話の送受信部に受信された後、相手から見た「相手の音声信号」として相手が用いる携帯電話の音声取込部に取り込まれる。本人の音声信号は、送受信部12から音声取込部13に取り込んでもよい。相手の音声信号は、相手が用いる携帯電話の内蔵マイクから相手が用いる携帯電話の送受信部に取り込まれて発信されて送受信部12に受信され、前記本人の音声信号同様、音声取込部13に取り込まれる。こうして取り込まれた本人及び相手の音声信号は、音声取込部13により、それぞれアバター動画生成部16へ送り込まれる。

【0048】

動画取込部14は、相手が用いる携帯電話の内蔵カメラから送受信部に取り込まれて発信され、送受信部12に受信される相手の実写動画を取り込む。従来のテレビ電話機能付き携帯電話は、送受信部が受信した相手の実写動画を直接画像表示部に送り込み、画面に表示させていた。これに対し、本発明に係る携帯電話1は、相手の実写動画を動画取込部14から動画合成部17へ送り込み、本人アバター51の動画と合成したリズム画像5として、相手53の実写動画を画像表示部へ送り込む。また、内蔵カメラ113から送受信部12に取り込まれた本人の実写動画は、前記送受信部12から発信され、相手が用いる携帯電話の送受信部に受信された後、相手から見た「相手の実写動画」として相手が用いる携帯電話の動画取込部に取り込まれる。

【0049】

信号取込部15は、テンキー114から本人の操作信号を取り込む。この操作信号は、本人アバター51に特定の動きからなる特別動作を実行させる命令である。前記特定の動きは、テンキー114の各ボタン毎に割り当てておき、本人が選択してボタンを押すことで、本人アバター51の特別動作として実行される。本人の操作信号は、テンキー114から送受信部12に取り込まれて発信され、相手が用いる携帯電話の送受信部に受信された後、相手から見た「相手の操作信号」として相手が用いる携帯電話の信号取込部に取り込まれる。本人の操作信号は、送受信部12から信号取込部15に取り込んでもよい。本人の操作信号は、信号取込部15によりアバター動画生成部16へ送り込まれる。ここで、本人による本人アバターの意図的な操作をさせない場合は、この信号取込部15を停止させるとよい。後掲の図3中、特別動作に係る流れは波線とし、信号取込部15の作動及び停止を選択的に切り換えることができることを表している。

【0050】

10

20

30

40

50

動画合成部17は、アバター動画生成部16から送り込まれた本人アバター51の動画と、動画取込部14から送り込まれた相手53の実写動画とを合成し、1枚のリズム画像5を生成する。ここで、各動画を「合成」するとは、例えば相手を奥側、本人アバターを手前側として配置し、本人アバターが被さる相手の部分は表示させないように、両動画を1枚のリズム画像にすることを意味し、1枚のリズム画像として見えれば各動画を異なるレイヤーにおける重ね合わせとしてもよい。こうして生成されたリズム画像5は、動画合成部17から画像表示部18へ送り込まれ、前記画像表示部18により携帯電話1の画面111に表示される。画像表示部18は、従来の携帯電話にも搭載される画像表示手段から構成できるが、既述したように、送受信部12から直接送り込まれた相手53の実写動画を画面111に表示するのではなく、動画合成部17で生成されたリズム画像5を画面111に表示する点が従来と異なる

10

【0051】

アバター動画生成部16は、音声取込部13から送り込まれた本人及び相手の音声信号から、また信号取込部15を作動させていれば信号取込部15から送り込まれた本人の操作信号から、本人アバター51が動くリズムタイミングを算出し、前記リズムタイミングで実行するリズム動作を作り出す。このアバター動画生成部16は、本人及び相手の音声信号と、本人の操作信号とを並列に取り込み、本人の音声信号から本人アバター51の話し手タイミングを推定し、相手の音声信号から本人アバター51の聞き手タイミングを推定し、そして本人の操作信号を取り込んだ時点を本人アバター51の特別タイミングとして、特に相互の競合を解消することなく、これらを重疊的に混在させてリズムタイミングを算出している（後掲図3参照）。

20

【0052】

そして、上記アバター動画生成部16は、推定された話し手タイミングに対して本人アバター51の話し手動作を、推定された聞き手タイミングに対して本人アバター51の聞き手動作を、決定された特別タイミングに対して本人アバター51の特別動作を作り出す。既述したように、リズムタイミングは話し手タイミング、聞き手タイミング、そして特別タイミングを重疊的に混在させているから、リズムタイミングで実行するリズム動作は、前記話し手動作、聞き手動作、そして特別動作が混在し、各タイミングによっては相互に組み合わさった複合的な動作となる。本人アバター51の動画は、前記リズムタイミングでリズム動作を実行する本人アバター51の動きとして生成される。

30

【0053】

次に、アバター動画生成部16における本人アバター51の聞き手タイミングの推定と聞き手動作の生成とを例にとり、具体的に説明する。図3はリズム画像5を生成する処理手順を表す処理フローチャート、図4は本人アバター51の話し手タイミングの推定と話し手動作の生成とを表す処理フローチャート、図5は本人アバター51の聞き手タイミングの推定と聞き手動作の生成とを表す処理フローチャート、図6は本人アバター51の特別タイミングの推定と特別動作の生成とを表す処理フローチャート、図7は聞き手タイミングのマクロ推定とミクロ推定とを表すタイミングチャートであり、図8は聞き手タイミングのマクロ推定とミクロ推定とを実行する論理回路のブロック図である。

40

【0054】

本発明を利用した携帯電話1は、図3に見られるように、通話開始により、「本人の音声信号取込」手順、「話し手タイミング推定」手順及び「話し手動作生成」手順の流れと、「相手の音声信号取込」手順、「聞き手タイミング推定」手順及び「聞き手動作生成」手順の流れと、そして「本人の操作信号取込」手順、「特別タイミング決定」手順及び「特別動作生成」手順とを並列に実行し、「本人アバターの動画生成」手順により、生成された話し手動作、聞き手動作及び特別動作からなるリズム動作に従って動く本人アバター51の動画を生成し、「相手の実写動画取込」手順により取り込まれた相手53の実写動画と前記本人アバター51の動画とを「リズム画像生成」手順により合成してリズム画像5を生成し、このリズム画像5を「リズム画像表示」手順により画面111に表示する一連の処理手順を繰り返す。前記処理手順は、通話終了になれば終了する。

50

【 0 0 5 5 】

話し手タイミングは、図4に見られるように、取り込まれた本人の音声信号について、「話し手マクロ推定値算出」手順及び「話し手ミクロ推定値算出」手順を同時に実行して逐次話し手マクロ推定値及び話し手ミクロ推定値を算出し、「話し手タイミング推定」手順によりマクロ推定及びミクロ推定が共に肯定された時点話し手タイミングとして推定して、話し手動作は、続く「話し手動作生成」手順により本人アバター51の口513の開閉動作と前記口513を除く頭511、目512、左前腕516、左上腕517、右前腕514、右上腕515及び胴518のランダムな身体動作とを組み合わせさせた話し手動作を生成する（可動部各部は図9参照、以下同じ）。

【 0 0 5 6 】

聞き手タイミングは、図5に見られるように、取り込まれた相手の音声信号について、「聞き手マクロ推定値算出」手順及び「聞き手ミクロ推定値算出」手順を同時に実行して逐次聞き手マクロ推定値及び聞き手ミクロ推定値を算出し、「聞き手タイミング推定」手順によりマクロ推定及びミクロ推定が共に肯定された時点聞き手タイミングとして推定して、聞き手動作は、続く「聞き手動作生成」手順により本人アバター51の頭511の頷き動作と前記頭511を除く目512、口513、左前腕516、左上腕517、右前腕514、右上腕515及び胴518のランダムな身体動作とを組み合わせさせた話し手動作を生成する。

【 0 0 5 7 】

特別タイミングは、図6に見られるように、本人の操作信号の存在を判断し、前記操作信号の存在が肯定されれば、「特別タイミング決定」手順により操作信号が取り込まれた時点、正確には前記操作信号の存在が肯定された時点特別タイミングとして決定し、特別動作は、続く「特別動作生成」手順により本人アバター51の頭511、目512、口513、左前腕516、左上腕517、右前腕514、右上腕515及び胴518を予め選択した身体動作として生成する。これら話し手動作、聞き手動作及び特別動作は、それぞれが生成された時点で各個に独立しているが、各動作に従って本人アバター51を動かすことにより、重畳的に足し合わされ、不規則なりズム動作を構成する。

【 0 0 5 8 】

本発明を特徴づける聞き手タイミングの推定について説明する。本発明は、会話全体から聞き手タイミングを推定するマクロ推定と、直近の会話から聞き手タイミングを推定するミクロ推定とによる二重推定を採用することにより、会話全体の身体リズムを損なわず、また直近の会話に関係づけた聞き手タイミングを推定する。まず、前記マクロ推定は、相手の音声信号をON/OFF信号とみなし、そして相手の音声信号を呼気段落区分に従って区分した各呼気段落に含まれるON信号の割合を移動平均して算出される聞き手マクロ推定値を、予め設定した聞き手マクロ閾値とを比較する手順で実行される。

【 0 0 5 9 】

人の音声信号（発話音声信号）は、図7に見られるように、文字を表す連続した波の部分と、呼吸による波の途切れた部分とに分けることができる。こうした音声信号を前記呼吸の部分で区切ることが「呼気段落区分」と言い、前記区切った単位を「呼気段落」と呼ぶ。ここで、音声信号を強度閾値（例えば環境ノイズより高い信号強度）でON信号（「1」）及びOFF信号（「0」）の2値化処理してON/OFF信号とみなせば、各呼気段落はON信号及びOFF信号の繰り返しからなる音声信号を含むことになる。この場合、各呼気段落におけるON信号は、相手との会話において本人が話し続けると割合が増え、逆に本人が黙り続けると割合が低下する。これから、会話全体の身体リズムは、前記呼気段落におけるON信号の割合の変化として表される。

【 0 0 6 0 】

そこで、本発明のマクロ推定は、下記数1及び数2に見られるように、各呼気段落におけるON信号の割合を移動平均した聞き手マクロ推定値 $E_{macro}(i)$ を算出し、前記聞き手マクロ推定値 $E_{macro}(i)$ が予め設定した聞き手マクロ推定閾値 $E_{macro-th}$ を超えた場合に、会話全体における聞き手タイミングの存在を推定する。ここで、 $a(j)$ は聞き手マクロ推定係数、 $R(i-j)$ は各呼気段落に含まれるON信号の割合（数2）、 $u(i)$ はノイズ、 $T(i-j)$ は各呼

10

20

30

40

50

気段落に含まれるON信号の時間、 $S(i-j)$ は各呼気段落に含まれるOFF信号の時間である。ノイズ $u(i)$ は、各呼気段落に含まれるON信号の割合 $R(i-j)$ 及び聞き手マクロ推定係数 $a(j)$ の積である移動平均に加えることで、数1より算出される聞き手マクロ推定値 $E_{macro}(i)$ に不規則性を加え、マクロ推定による聞き手タイミングの存在の判断がばらつくようにしている。

【0061】

【数1】

$$E_{macro}(i) = \sum_{j=1}^J a(j)R(i-j) + u(i)$$

10

【0062】

【数2】

$$R(i-j) = \frac{T(i-j)}{T(i-j) + S(i-j)}$$

【0063】

上記マクロ推定により、会話全体から聞き手タイミングが推定されても、前記聞き手タイミングの推定は具体的な時点を明確にしない。そこで、本発明では、ミクロ推定により、直近の身体リズムに対応した聞き手タイミングを推定する。このミクロ推定では、現在の音声信号に対する過去の音声信号の関わり合いの濃淡が身体リズムを表しているとして、予め設定したサンプリング間隔でサンプリングした音声信号の信号強度を移動平均して、聞き手タイミングを推定している。ここで、直近の身体リズムを反映するミクロ推定のみで、聞き手タイミングを導き出すことも考えられるが、単純にサンプリングした音声信号強度の移動平均から算出される聞き手ミクロ推定値 $E_{micro}(i)$ は、比較的容易に聞き手ミクロ推定閾値 $E_{micro-th}$ を超えやすく、適切な聞き手ミクロ推定閾値 $E_{micro-th}$ の設定が難しい。このため、本発明ではマクロ推定とミクロ推定とを掛け合わせることで、適度な頻度で聞き手タイミングが推定されるようにしている。

20

【0064】

具体的には、数3により、上述同様な移動平均の結果として聞き手ミクロ推定値 $E_{micro}(i)$ を算出し、この聞き手ミクロ推定値 $E_{micro}(i)$ が予め設定した聞き手ミクロ推定閾値 $E_{micro-th}$ を超えるか否かにより、聞き手タイミングを推定する。ここで、 $b(k)$ は聞き手ミクロ推定係数、 $V(i-k)$ はサンプリングした音声信号強度、 $w(i)$ はノイズである。ノイズ $w(i)$ は、サンプリングした音声信号強度 $V(i-k)$ 及び聞き手ミクロ推定係数 $b(k)$ の積である移動平均に加えることで、数3より算出される聞き手ミクロ推定値 $E_{micro}(i)$ に不規則性を加え、ミクロ推定による聞き手タイミングの存在の判断がばらつくようにしている。

30

【0065】

【数3】

$$E_{micro}(i) = \sum_{k=1}^K b(k)V(i-k) + w(i)$$

【0066】

こうしたマクロ推定とミクロ推定とを掛け合わせて聞き手タイミングを推定する論理回路は、図8に見られるように構成できる。この論理回路は、マクロ推定がスイッチとなり、マクロ推定が肯定された場合、すなわち前記スイッチがONになった場合のみ、ミクロ推定により具体的な聞き手タイミングを推定される構成になっている。このように、マクロ推定及びミクロ推定は、論理的にはマクロ推定が上位にあり、ミクロ推定が下位にあって、マクロ推定が先行して判断されるが、マクロ推定及びミクロ推定はいずれも取り込まれた相手の音声信号から聞き手マクロ推定値及び聞き手ミクロ推定値を算出することから、実際には前記聞き手マクロ推定値及び聞き手ミクロ推定値を並列に算出しておき、両推定がいずれも肯定された時点を、聞き手タイミングとして推定する(図5参照)。

40

【0067】

50

このほか、本例では、本人アバター51の聞き手動作に多様性を付加するため、図8に見られるように、ミクロ推定に際して算出した聞き手ミクロ推定値 $E_{micro}(i)$ に対し、頭511の頷き動作の聞き手タイミングを判断する聞き手ミクロ推定閾値 $E_{micro-th}$ と、頭511を除くその他の可動部をランダムに選択した身体動作の聞き手タイミングを判断する聞き手ミクロ推定閾値 $E_{micro-th}$ とを異ならせ、後者の聞き手ミクロ推定閾値 $E_{micro-th}$ を低くすることで身体動作の実行頻度を頭の頷き動作の実行頻度より多くしている。また、目512の瞬き動作は、頭511の頷き動作の聞き手タイミングから指数分布させた聞き手タイミングを用い、頭511の頷き動作と目512の瞬き動作とを関連付けながら、目512の瞬き動作に不規則性を付加している。聞き手動作は、聞き手タイミングを基準とした時系列に沿って実行されればよいから、具体的な聞き手動作は自由に設定できる。

10

【0068】

本人アバター51の話し手タイミングも、上述の聞き手タイミング同様、本人の音声信号に基づくマクロ推定とミクロ推定とを掛け合わせ、両推定が肯定された時点として推定することができるが、話し手動作は聞き手動作と異なり、煩雑になることで困る動きが少ないため、例えばミクロ推定のみで話し手タイミングを推定してもよい。また、マクロ推定及びミクロ推定を掛け合わせた場合や、ミクロ推定のみの場合のいずれでも、話し手動作のうち、話し手アバターの可動部のいずれかの動作に対する話し手ミクロ推定閾値 $E_{micro-th}$ を、他の動作に対して高く又は低く設定することにより、上述同様、実行頻度を異ならせることができる。そして、口513の開閉動作は、本人の音声信号強度に比例した単純な開閉運動とすればよい。その他子細については、上述した聞き手タイミングの推定及び聞き手動作の生成と同様であるため、説明は省略する。

20

【0069】

最後に、リズムタイミングでリズム動作を実行して動く本人アバター51の動画について説明する。図9は本人アバター51の正面図、図10は話し手タイミング、聞き手タイミング及び特別タイミングからなるリズムタイミングを表すタイミングチャート、図11は本人アバター51の右前腕514の基本的な動きを表す部分正面図、図12は本人アバター51の右前腕514の動きの重なりを表す部分正面図であり、図13は画面111の表示例を表す本人が用いる携帯電話1の部分正面図である。

【0070】

本例の本人アバター51は、図9に見られるように、頭511、目512、口513、右前腕514、右上腕515、左前腕516、左上腕517、胴518を可動部とし、各可動部を次のように動かす。まず、左上腕517及び右上腕515は胴518に対して所定角度で上下に往復運動（回転運動）させ、左前腕516は左上腕517に対し、右前腕514は右上腕515に対して所定角度で上下に往復運動（回転運動）させる。胴518は所定角度で左右に往復運動（回転運動）させ、頭511は胴518に対して所定角度で前後に往復運動（回転運動）させる（頭511の頷き動作）。そして、目512及び口513は開閉運動させる（目512の瞬き動作、口513の開閉動作）。各可動部は、それぞれの運動時間を異ならせたり、運動時間を等しくしながら各運動の運動範囲を異ならせることにより、動きに変化をもたせることができる。

30

【0071】

話し手動作は、口513の開閉動作と前記口513を除くその他の可動部のランダムな身体動作との組み合わせとして、聞き手動作は頭511の頷き動作と前記頭511を除くその他の可動部をランダムに選択した身体動作との組み合わせとして、そして特別動作は可動部を予め選択した身体動作から、それぞれ構成する。特別動作は、本人の意図的な操作により本人アバター51を動かすものであるから、例えば操作信号の内容に合わせて手の表示を適宜変更する等、特別な可動部を設定してもよい。また、話し手動作又は聞き手動作は、各動作の実行開始時点における各可動部の現在位置（後述の中間位置）から所定角度だけ運動させるが、特別動作は独立した意味合いを有するため、あくまで各可動部に予め設定した変位位置まで運動させる。すなわち、特別動作は、可動部の現在位置と変位位置との差分だけ運動させる。

40

【0072】

50

本人アバター51のリズムタイミングは、話し手タイミング、聞き手タイミング、そして特別タイミングの重畳的な重ね合わせであり、図10に見られるような関係にある。説明の便宜上、ここでは各可動部の話し手マイクロ推定閾値Emicro-th及び聞き手マイクロ推定閾値Emicro-thはすべて同じとする。すなわち、話し手動作は口513の開閉動作と前記口513を除くその他の可動部のランダムな身体動作とを同時に実行し、聞き手動作は頭511の頷き動作と前記頭511を除くその他の可動部をランダムに選択した身体動作とを同時に実行させる。

【0073】

まず、話し手タイミングTM1が推定されると、本人アバター51は話し手動作を実行し、前記話し手動作が終了した時点で推定された聞き手タイミングHM1に従って、今度は聞き手動作を実行する。ここまでは、話し手タイミング、聞き手タイミング及び特別タイミング同士（リズムタイミング同士）の競合や、話し手動作、聞き手動作又は特別動作（リズム動作）に対する話し手タイミング、聞き手タイミング又は特別タイミング（リズムタイミング）の競合はないため、リズムタイミングでのリズム動作は、それぞれ単一の話し手動作及び聞き手動作でしかない。例えば、話し手動作又は聞き手動作で右前腕514が選択されると、この右前腕514は、図11に見られるように、原点位置を基準位置として前記基準位置から所定角度離れた変位位置まで回転運動し、再び原点位置に戻る。

【0074】

しかし、続く話し手タイミングTM2に従った話し手動作の実行中、話し手タイミングTM3が推定されると、話し手タイミングTM2で実行する話し手動作に対して話し手タイミングTM3が競合する。本例では、前記競合を特に解消することなく、話し手タイミングTM2で実行中の話し手動作に、話し手タイミングTM3で新たな話し手動作に実行して重ね合わせる。すなわち、話し手タイミングTM2及び話し手タイミングTM3いずれの話し手動作でも右前腕514が選択されると、この右前腕514は、図12に見られるように、まず話し手タイミングTM2で実行する話し手動作により、原点位置を基準位置1として変位位置1まで回転運動し、今度は基準位置1に戻る途中の中間位置を話し手タイミングTM3で実行する話し手動作の基準位置2として、前記基準位置2から変位位置2まで回転運動して、その後原点位置に戻る。

【0075】

ここで、単一の話し手動作における右前腕514の動き（図11参照）と、複合した話し手動作における右前腕514の動き（図12参照）とを比較すれば明らかなように、複数の話し手動作が重なり合う場合の方が、可動部の動きが大きくなる。また、リズムタイミングの重なり合いは全くの任意であるため、重なり合う話し手動作の基準位置（上記基準位置2）は、重なり合うリズムタイミングによって異なることになり、結果として得られる複合的なリズム動作はランダムなものとなる。これにより、本人アバター51の動きから規則性がなくなり、多様性が増えることになり、より自然なリズム動作を実現できる。

【0076】

しかし、話し手動作と聞き手動作とは厳密には意味合いの異なる動きであり、話し手タイミング及び聞き手タイミングはそれぞれ並列に推定することから、前記両タイミングが競合して、話し手動作と聞き手動作の重なり合いがうまくいかず、不自然なリズム動作が作り出される可能性がある。また、特別動作は単独で本人の意図を反映した動きを作り出すものだから、他のリズム動作が重なり合わない方が好ましい場合も考えられる。このようなリズムタイミング同士の競合や、リズム動作に対するリズムタイミングの競合を解消する場合、次のようにするとよい。

【0077】

まず、特別動作がない場合、次のように話し手タイミング及び聞き手タイミングの競合を解消する。具体的には、実行中のリズム動作に話し手動作又は聞き手動作が重なり合うことに問題はないため、

<条件1> 話し手タイミングと聞き手タイミングとが競合しなければ、

<対応1> 先行する話し手タイミング及び聞き手タイミングに従って、話し手動作又は聞

10

20

30

40

50

き手動作を作り出して実行するが、

<条件2>話し手タイミングと本人アバターの聞き手タイミングとが競合すれば、

<対応2>優先度の高い話し手タイミング又は聞き手タイミングを選択し、話し手動作又は聞き手動作を作り出して実行する。

<対応2>における優先度は固定して常に話し手タイミング又は聞き手タイミングのいずれかを選択させてもよいが、例えば実行中又は直前に実行を終了したリズム動作が話し手動作であれば話し手タイミングを優先させ、逆に実行中又は直前に実行を終了したリズム動作が聞き手動作であれば聞き手タイミングを優先させてもよい。

【0078】

次に、特別動作がある場合、特別動作の取り扱いの特殊性から、次のように話し手タイミング、聞き手タイミング及び特別タイミングの競合を解消する。まず、

<前提1>リズム動作が実行中でないか、話し手動作又は聞き手動作をリズム動作として実行中で、

<条件1-1>話し手タイミング、聞き手タイミング又は本人アバターの特別タイミングが競合しなければ、

<対応1-1>先行する話し手タイミング、聞き手タイミング及び特別タイミングに従って、話し手動作、聞き手動作又は特別動作を作り出して実行する。

しかし、<前提1>において

<条件1-2>話し手タイミング又は聞き手タイミングいずれかと特別タイミングとが競合すれば、

<対応1-2>特別タイミングを優先させて特別動作を作り出し、実行する。

また、同じく前提1>において

<条件1-3>話し手タイミング又は聞き手タイミングのみが競合すれば、

<対応1-3>優先度の高い話し手タイミング又は聞き手タイミングを選択し、話し手動作又は聞き手動作を作り出して実行する。

ここで、<対応1-3>における優先度は、既述したように固定していてもよいし、変動させるものでもよい。

【0079】

上述に対し、

<前提2>特別動作をリズム動作として実行中で、

<条件2-1>話し手タイミング、聞き手タイミング又は本人アバターの特別タイミングが競合しなければ、

<対応2-1>先行する話し手タイミング、聞き手タイミング又は特別タイミングを選択し、リズム動作の実行終了時点話し手タイミング、聞き手タイミング又は特別タイミングとして、話し手動作、聞き手動作又は特別動作を作り出して実行する。

また、前記<前提2>において

<条件2-2>話し手タイミング又は聞き手タイミングいずれかと特別タイミングとが競合すれば、

<対応2-2>特別タイミングを選択し、リズム動作の実行終了時点特別タイミングとして、特別動作を作り出して実行する。

更に、同じく前提2>において

<条件2-3>話し手タイミング又は聞き手タイミングのみが競合すれば、

<対応2-3>優先度の高い話し手タイミング又は聞き手タイミングを選択し、リズム動作の実行終了時点を選択した話し手タイミング又は聞き手タイミングとして、話し手動作又は聞き手動作を作り出して実行する。

ここで、<対応2-3>における優先度は、既述したように固定していてもよいし、変動させるものでもよい。この<前提2>における競合の解消は、リズム動作として実行中の特別動作を阻害しないように、続く話し手タイミング、聞き手タイミング又は特別タイミングをリズム動作の実行終了時点まで遅延させる点に特徴がある。

【0080】

10

20

30

40

50

こうして、話し手動作、聞き手動作及び特別動作からなるリズム動作を実行させて動く本人アバター51の動画は、必要により上記競合の解消を図って、アバター動画生成部16により生成される。そして、動画合成部17により前記本人アバター51の動画と相手53の実写動画と合成されてリズム画像5となる。ここで、本例の携帯電話1は、同一画面111に表示される本人アバター51と相手53とが疑似的に会話していると本人に感得させることにより、より多くのノンバーバル情報を本人に取得させる。これから、具体的なリズム画像5は、図13に見られるように、本人アバター51は顔が見える程度の斜視に後ろ向きで画面右下に配し、相手53は前向きで画面中央に配すると両者の対面関係が明らかとなり、疑似的な会話をしていると本人感得させやすくなる。こうした本人アバター51と相手53との位置関係や向き、更に大小関係等は、動画合成部17により調整するとよい。このほか、動画合成部17は、リズム画像5の背景等を本人アバター51の動画と相手53の実写動画に重ね合わせる。

10

【0081】

以下、本発明の応用例である携帯電話やその応用となる会話支援端末について、それぞれ上記例(図1 - 図13参照)と異なる点について、説明する。図14は特別動作を含むリズム動作に従って動く相手アバター52の動画からなるリズム画像5を用いた携帯電話2の使用状態を示す斜視図、図15は特別動作を含むリズム動作に従って動く相手アバター52の動画からなるリズム画像5を用いた携帯電話2の本発明に係る構成を示すブロック図、図16はリズム画像5を生成する処理手順を表す処理フローチャート、図17は画面211の表示例を表す本人が用いる携帯電話2の部分正面図であり、図18は本例の携帯電話2を応用した会話支援端末3の本発明に係る構成を示すブロック図である。

20

【0082】

この応用例は、話し手動作、聞き手動作及び特別動作を混在させて動く相手アバター52の動画からなるリズム画像5を用いた携帯電話2の例である。これから、図14に見られるように、本人及び相手が用いる各携帯電話2は、互いに本人及び相手の音声信号を送受信するほか、互いに本人及び相手の操作信号を送受信する。本人の音声信号は、相手から見れば相手の音声信号、本人の操作信号は相手から見れば相手の操作信号になる。本人及び相手が用いる携帯電話2は同じである。ここで、本例の携帯電話2は、上記例(図1 ~ 図13参照)と異なり、本人及び相手の実写動画を送受信しない。これは、携帯電話2の回線容量が小さくても、また本人及び相手が用いる携帯電話2にカメラがなくても、本発明が利用

30

【0083】

携帯電話2は、図15に見られるように、音声取込部23、信号取込部25、アバター動画生成部26、そして画像表示部28から、本発明に係る主要部を構成する。この主要部は、従来より携帯電話に搭載されている情報処理装置をプログラムにより制御し、ソフト的に構成できる。ここで、例えば音声取込部23、信号取込部25及び画像表示部28は、処理内容が特定されているので、専用LSI等を用いてハード的に構成してもよい。ここで、既述したように、本例の携帯電話2はカメラを必要とせず、また上記例に比べて動画合成部が省略されていることから、本発明を利用するための仕様要求が低くなっている。これから、主要部はソフト的に構成する方が、その後の仕様変更に応じやすくなる。

40

【0084】

音声取込部23は内蔵マイク212から本人の音声、相手が用いる携帯電話から送受信部22を通じて相手の音声信号を取り込み、それぞれアバター動画生成部26へ送り込む。また、画像表示部28は、アバター動画生成部26で生成された相手アバター52の動画、すなわちリズム画像5を取り込み、画面211に表示する。この音声取込部23及び画像表示部28の詳細は、上記例と同じであるから、説明は省略する。

【0085】

信号取込部25は、相手が用いる携帯電話のテンキーから相手が用いる携帯電話の送受信部に取り込まれて発信され、送受信部22に受信された相手の操作信号を取り込む。この操作信号は、相手アバター52に特定の動きからなる特別動作を実行させる命令である。各特別

50

動作は、テンキーの各ボタン毎に割り当てておき、相手が選択してボタンを押すことで、相手アバターの特別動作として実行される。本人の操作信号は、テンキー214から送受信部22に取り込まれて発信され、相手が用いる携帯電話の送受信部に受信された後、相手から見た「相手の操作信号」として相手が用いる携帯電話の信号取込部に取り込まれる。相手の操作信号は、信号取込部25によりアバター動画生成部26へ送り込まれる。

【0086】

アバター動画生成部26は、音声取込部23から送り込まれた本人及び相手の音声信号から、また信号取込部25から送り込まれた相手の操作信号から、相手アバター52が動くリズムタイミングを算出し、前記リズムタイミングで実行するリズム動作を作り出し、前記リズム動作に従って動く相手アバター52の動画そのものをリズム画像5として生成する。このアバター動画生成部26は、図16に見られるように、本人及び相手の音声信号と、相手の操作信号とを並列に取り込み、相手の音声信号から相手アバターの話し手タイミングを推定し、本人の音声信号から相手アバターの聞き手タイミングを推定し、そして相手の操作信号を取り込んだ時点を相手アバターの特別タイミングとして、特に相互の競合を解消することなく、これらを重疊的に混在させてリズムタイミングを算出する。

10

【0087】

そして、上記アバター動画生成部26は、推定された話し手タイミングに対して相手アバター52の話し手動作を、推定された聞き手タイミングに対して相手アバター52の聞き手動作を、決定された特別タイミングに対して相手アバター52の特別動作を作り出す。既述したように、リズムタイミングは話し手タイミング、聞き手タイミング、そして特別タイミングを重疊的に混在させているから、リズムタイミングで実行するリズム動作は、前記話し手動作、聞き手動作、そして特別動作が混在し、各タイミングによっては相互に組み合わせられた複合的な動作となる。相手アバター52の動画、すなわちリズム画像5は、前記リズムタイミングでリズム動作を実行する相手アバター52の動きとして生成される。そして、このアバター動画生成部26は、背景等を相手アバター52の動画に重ね合わせてリズム画像5を生成する。

20

【0088】

本例の携帯電話2は、身体リズムに従って動く相手アバター52が相手の意図的な操作を表す特別動作で動くことにより、本人に対する相手アバター52の関わり合いが緊密であるとして本人に身体リズムを感得させ、この本人にノンバーバル情報を取得させる。これから、具体的なリズム画像5は、図17に見られるように、相手アバター52は前向きで画面中央に配し、本人と相手アバター52とが向かい合う対面関係にするとよい。この場合、相手アバター52の向き及び姿勢は、アバター動画生成部26により調整できるため、相手アバター52は正しく本人に対面する向き及び姿勢で画面211に表示させることができる。

30

【0089】

本例に示すリズム画像5は、上記例に示すリズム画像(図13)に実写動画として表示される相手と同じく、特別動作として「ピースサイン」を実行させている。これは、相手が「ピースサイン」の特別動作を割り当てたテンキーを押し、意図的に相手アバター52に「ピースサイン」を実行させた場合に当たる。このほか、図示は省略するが、強い肯定反応を表すゆっくり又は早くて深い頭の頷き動作や、強い否定反応を表すゆっくり又は早くて大きな頭の(左右)振り動作等が、本発明における話し手動作及び聞き手動作に混在させる特別動作として有用であり、本人に身体リズムを感得させやすくなる。実写動画における相手は、相手が用いる携帯電話のカメラとの位置関係により、必ずしも相手と正しく対面しない。しかし、本例の相手アバター52は正しく本人と対面させることができる(図13及び図17を比較対照)。これにより、本人は相手アバター52の関わり合いが緊密であると感じやすくなり、この相手アバター52からなるリズム画像5を見る本人に身体リズムを感得させ、この本人にノンバーバル情報を取得させることができる。

40

【0090】

応用例の携帯電話2を利用すれば、図18に見られるように、会話支援端末3を構成できる。この会話支援端末3は、上記携帯電話2(図14~図17参照)から送受信部を除き、音声

50

取込部33、信号取込部35、アバター動画生成部36、そして画像表示部38から、本発明に係る主要部を構成する。この主要部は、会話支援端末3に搭載される情報処理装置をプログラムにより制御してソフト的に構成できるほか、例えば音声取込部33、信号取込部35及び画像表示部38は、専用LSI等を用いてハード的に構成してもよい。

【0091】

音声取込部33は、内蔵マイク312から本人及び相手の音声信号を取り込み、前記本人及び相手の音声信号をそれぞれアバター動画生成部36へ送り込む。信号取込部35は、データグループ314から本人の操作信号を取り込む。この操作信号は、本人アバターに特定の動きからなる特別動作を実行させる命令である。特別動作は、データグループ314の各指の動きに応じた可動部を動かすことにより、本人アバターの特別動作として実行される。本人の操作信号は、信号取込部35によりアバター動画生成部36へ送り込まれる。

10

【0092】

アバター動画生成部36は、音声取込部33から送り込まれた本人及び相手の音声信号から、また信号取込部35から送り込まれた本人の操作信号から、本人アバターが動くリズムタイミングを算出し、前記リズムタイミングで実行するリズム動作を作り出し、このリズム動作に従って動く本人アバターの動画を生成する。この本人アバターの動画がリズム画像であり、前記アバター動画生成部36により画面表示部38へ送り込まれ、この画面表示部38により画面311に表示される。アバター動画生成部36は、背景等を本人アバター51の動画に重ねてリズム画像5を生成する。

【0093】

この会話支援端末3は、本人アバターの動きがはっきり分かるように、本人アバターの向き及び姿勢をアバター動画生成部により調整して、正しく本人及び相手に対面する向き及び姿勢で本人アバターを画面311(図17中画面211に相当)に表示させるとよい。これにより、本人の意図的な操作を表す特別動作を加えたりリズム動作に従って動く本人アバターを見る相手に身体リズムを感得させ、更には前記特別動作により本人の意図を相手に伝達することで、十分なノンバーバル情報を相手に与えることができる。また、本人が会話の不自由な身体障害者である場合、本人アバターが本人に身体リズムを感得させ、ノンバーバル情報を取得させることができる。こうして、この会話支援端末3は、特に健常者と会話する身体障害者の前記健常者に対する意思表示又は感情表現を補い、健常者と身体障害者との会話を円滑にする。

20

30

【0094】

図19は本人アバター51及び相手アバター52の動画を組み合わせたりリズム画像5を用いた携帯電話4の使用状態を示す斜視図、図20は本人アバター51及び相手アバター52の動画を組み合わせたりリズム画像5を用いた携帯電話4の本発明に係る構成を示すブロック図、図21はリズム画像5を生成する処理手順を表す処理フローチャートであり、図22は画面411の表示例を表す本人が用いる携帯電話4の部分正面図である。本例の携帯電話4は、上記例(図1-図13)が用いた相手の実写動画に代えて、相手アバター52を用いた構成であり、携帯電話4の回線容量が小さくても、また本人及び相手を用いる携帯電話4にカメラがなくても利用できる利点がある。

【0095】

この応用例は、話し手動作及び聞き手動作を混在させて動く本人アバター51の動画と話し手動作及び聞き手動作を混在させて動く相手アバター52の動画とを重ね合わせたリズム画像5を用いた携帯電話4の例である。これから、図19に見られるように、本人及び相手を用いる各携帯電話4は、互いに本人及び相手の音声信号及び操作信号を送受信する。ここで、本人の音声信号は相手から見れば相手の音声信号、本人の操作信号は相手から見れば相手の操作信号になる。

40

【0096】

携帯電話4は、図20に見られるように、音声取込部43、信号取込部45、アバター動画生成部46、動画合成部47、そして画像表示部48から、本発明に係る主要部を構成する。この主要部は、従来より携帯電話に搭載されている情報処理装置をプログラムにより制御し、ソ

50

フト的に構成できる。ここで、例えば音声取込部43、信号取込部45、動画合成部47及び画像表示部48は、処理内容が特定されているので、専用LSI等を用いてハード的に構成してもよい。

【0097】

音声取込部43は、内蔵マイク412から本人の音声信号を取り込む。前記本人の音声信号は、内蔵マイク412から送受信部42に取り込まれて発信され、相手が用いる携帯電話の送受信部42に受信された後、相手から見た「相手の音声信号」として相手が用いる携帯電話の音声取込部に取り込まれる。本人の音声信号は、送受信部42から音声取込部43に取り込んでもよい。また、相手の音声信号は、相手が用いる携帯電話の内蔵マイクから相手が用いる携帯電話の送受信部に取り込まれて発信されて送受信部42に受信され、前記本人の音声信号同様、音声取込部43に取り込まれる。こうして取り込まれた本人及び相手の音声信号は、音声取込部43により、それぞれアバター動画生成部46へ送り込まれる。

10

【0098】

信号取込部45は、テンキー414から本人の操作信号を取り込む。この操作信号は、本人アバター51に特定の動きからなる特別動作を実行させる命令である。特別動作は、テンキーの各ボタン毎に割り当てておき、本人が選択してボタンを押すことで、本人アバター51の特別動作として実行される。本人の操作信号は、送受信部42から信号取込部45に取り込んでもよい。また、相手の操作信号は、相手が用いる携帯電話のテンキーから相手が用いる携帯電話の送受信部に取り込まれて発信されて送受信部42に受信され、前記本人の操作信号同様、信号取込部45に取り込まれる。こうして取り込まれた本人及び相手の操作信号は、信号取込部45によりアバター動画生成部46へ送り込まれる。ここで、本人による本人アバター51の意図的な操作や、相手による相手アバター52の意図的な操作をさせない場合は、本人又は相手が用いる携帯電話4の各信号取込部45を停止させるとよい。

20

【0099】

アバター動画生成部46は、音声取込部43から送り込まれた本人及び相手の音声信号から、また信号取込部45から送り込まれた本人及び相手の操作信号から、本人アバター51及び相手アバター52それぞれのリズムタイミングを算出し、前記リズムタイミングで実行する各リズム動作を作り出す。このアバター動画生成部46は、図21に見られるように、本人アバター51に対する本人及び相手の音声信号、本人の操作信号、相手アバター52に対する本人及び相手の音声信号、そして相手の操作信号を並列に取り込む。そして、本人の音声信号から話し手タイミングを推定し、相手の音声信号から聞き手タイミングを推定し、そして本人の操作信号を取り込んだ時点を実特別タイミングとして、本人アバター51のリズムタイミングを算出し、同様に本人の音声信号から聞き手タイミングを推定し、相手の音声信号から話し手タイミングを推定し、そして相手の操作信号を取り込んだ時点を実特別タイミングとして、相手アバター52のリズムタイミングを算出している。

30

【0100】

そして、上記アバター動画生成部46は、本人アバター51の話し手動作、聞き手動作及び特別動作を作り出し、同様に相手アバター52の話し手動作、聞き手動作及び特別動作を作り出す。既述したように、リズムタイミングは話し手タイミング、聞き手タイミング、そして特別タイミングを重畳的に混在させているから、リズムタイミングで実行するリズム動作は、前記話し手動作、聞き手動作、そして特別動作が混在し、各タイミングによっては相互に組み合わせさせた複合的な動作となる。本人アバター51及び相手アバター52の各動画は、前記リズムタイミングでリズム動作を実行する本人アバター51及び相手アバター52の動きとして生成される。

40

【0101】

動画合成部47は、アバター動画生成部46から送り込まれた本人アバター51及び相手アバター52の各動画を合成し、1枚のリズム画像5を生成する。ここで、本人アバター51及び相手アバター52の各動画は、例えば相手アバター52を奥側、本人アバター51を手前側として配置し、本人アバター51が被さる相手アバター52の部分は表示させないように、両動画を1枚のリズム画像5にする。各動画は、1枚のリズム画像5として見えれば、それぞれ異

50

なるレイヤーにおける重ね合わせとしてもよい。こうして生成されたリズム画像5は、動画合成部47から画像表示部48へ送り込まれ、前記画像表示部48により携帯電話4の画面411に表示される。動画合成部47は、本人アバター51の動画と相手アバター52の動画とに、背景等を重ね合わせてリズム画像5を生成する。

【0102】

ここで、本例の携帯電話4は、同一画面411に表示される本人アバター51と相手アバター52とが疑似的に会話していると本人に感得させると共に、本人アバター51及び相手アバター52がそれぞれのリズムタイミングに従って、リズム動作、例えば話し手動作及び聞き手動作を実行することにより、より多くのノンバーバル情報を本人に取得させる。これから、具体的なリズム画像5は、図22に見られるように、本人アバター51は顔が見える程度の斜視に後ろ向きで画面右下に配し、相手は前向きで画面中央に配すると両者の対面関係が明らかとなり、疑似的な会話をしていると本人に感得させやすくなる。本例では、相手アバター52が「ピースサイン」からなる特別動作を実行している。また、相手アバター52は、画面411を見る本人と対面させることにより、相手アバター52との関わり合いが緊密であると本人に感じさせやすくなり、このリズム画像5を見る本人に身体リズムを感得させ、ノンバーバル情報を取得させやすくすることもできる。

【実施例1】

【0103】

本発明の携帯電話に用いるリズム画像が、本人に身体リズムを感得させ、ノンバーバル情報をどの程度伝達しているかを確認するため、コンピュータにより生成したリズム画像5を、本人及び相手となる各モニター（被験者）に見せながら会話してもらい、前記リズム画像5を用いた会話について官能評価してもらった試験を実施した。モニターは、18歳 - 24歳の男女20人ずつ、計40人で、2人ずつを組として試験に参加させ、各人毎に官能評価させた。

【0104】

本人及び相手となる各被験者は、直接会話できない個別の部屋に分かれてもらい、図23に見られるように、E-thernetケーブルでLAN接続された小型ノートパソコン6（SONY製 VAI O TYPE U VGN-U50）をそれぞれに渡し、前記小型ノートパソコン6の画面611に本発明のリズム画像5（実施例1）のほか、比較画像（比較例1、比較例2）を表示した状態で、各小型ノートパソコン6に接続したマイク付きヘッドフォン612を使って会話してもらった。

【0105】

実施例1は、話し手動作及び聞き手動作を混在させて動く本人アバター51の動画と話し手動作及び聞き手動作を混在させて動く相手アバター52の動画とを重ね合わせたリズム画像5に、特別動作を追加している。特別動作は、本人アバター51及び相手アバター52の各可動部を単独で意図的に動かす簡略なものとした。本人アバター51及び相手アバター52は、それぞれ座った姿勢の熊のぬいぐるみを模したキャラクタで、可動部として頭、口、目、左前腕、左上腕、右前腕、右上腕及び胴（腰）を設定している。リズム画像は、本人アバター51及び相手アバター52が画面611の左右に分かれ、それぞれが内向きに斜め前方を向いて座っている構成とした。説明の便宜上、図23中、左側のモニターを本人、右側のモニターを相手としている。

【0106】

上記可動部は、次のように動作する。頭は、前後に0.3radの範囲で1.2rad/secの角速度で回転運動する（頭の傾き動作）ほか、左右に0.4radの範囲で3.6rad/secの角速度で回転運動する。左上腕及び右上腕は、胴に対して上下に0.8radの範囲で4.0rad/secの角速度で回転運動し、左前腕は前記左上腕に対し、右前腕は前記右上腕に対して、それぞれ上下に0.8radの範囲で4.0rad/secの角速度で回転運動する。そして、胴（腰）は前方に0.1radの範囲、後方に0.05radの範囲で、それぞれ0.4rad/secの角速度で回転運動する。前記胴の回転運動は、頭及び腕を含む上半身の前後運動に相当する。このほか、口の開閉動作は、話し手動作に付随して音声信号の強弱に応じた動きとして、また目の開閉動作は、話し手タ

10

20

30

40

50

イミング又は聞き手タイミングから指数分布させた動きとして、それぞれ実行する。

【0107】

本人アバター51又は相手アバター52の特別動作は、各小型ノートパソコン6に接続した携帯電話型キーボード613(Mevael製 Keiboard

PA-101)を用いた。具体的なキー割当は、図24に見られるように、テンキーの上段に左から順に並ぶ「1」ボタンに頭の左への回転運動、「2」ボタンに頭の前後(縦)の回転運動、「3」ボタンに頭の右への回転運動、テンキーの中段に左から順に並ぶ「4」ボタンに右上腕の上下の回転運動、「5」ボタンに胸(腰)の前への回転運動、「6」ボタンに左上腕の上下の回転運動、そしてテンキーの下段に左から順に並ぶ「7」ボタンに右前腕の上下の回転運動、「8」ボタンに胸(腰)の後への回転運動、「9」ボタンに左前腕の上下の回転運動である。ボタン配置と、右上腕、左上腕、右前腕及び左前腕との関係を逆にしたのは、画面611に表示される本人アバター51及び相手アバター52の左右と本人又は相手の左右とが鏡面对称で逆になっているので、画面611に表示される本人アバター51及び相手アバター52を見たままに操作させるための便宜である。

10

【0108】

比較例1は、実施例1のリズム画像における本人アバター及び相手アバターからそれぞれ特別動作を取り除いた画像で、前記本人アバター及び相手アバターは話し手動作及び聞き手動作からなるリズム動作だけを実行する。比較例2は、実施例1のリズム画像における本人アバター及び相手アバターからそれぞれリズム動作を取り除いた画像で、前記本人アバター及び相手アバターは特別動作だけを実行する。ここで、比較例2において、口の開閉動作や目の瞬き動作がなくなると不自然なので、口の開閉動作及び目の瞬き動作のみはリズムタイミングに基づいて実行させている。

20

【0109】

各モニターには、実施例1、比較例1及び比較例2について、「7段階評価」及び「一対評価」をしてもらった。「7段階評価」では、本人及び相手が用いる小型ノートパソコン6の画面611に同じ実施例1、比較例1又は比較例2を30秒間ずつ表示させた状態で会話をしてもらい、会話や本人アバター51又は相手アバター52に対する「楽しさ」、「対話しやすさ」、「キャラクタになった感覚」、そして「キャラクタを操作している感覚」の評価項目について官能評価してもらった。評価段階は、普通の評価を「0」として、よい評価を「+1」、「+2」、そして「+3」の3段階、悪い評価を「-1」、「-2」、そして「-3」の3段階である。実施例1、比較例1及び比較例2の表示順は、予断が入らないように、ランダムにしている。

30

【0110】

「一対評価」では、本人及び相手が用いる小型ノートパソコン6の画面611に同じ実施例1、比較例1又は比較例2を30秒間ずつ表示させた状態で会話をしてもらい、先に表示させた実施例1、比較例1又は比較例2と、次に表示させた実施例1、比較例1又は比較例2とを対比させ、どちらがより好ましいかを総合的に比較評価してもらった。実施例1、比較例1及び比較例2の表示順は、予断が入らないように、ランダムにしている。

【0111】

実施例1における7段階評価の結果について、図25に示す。図25は、7段階評価における各官能評価項目についての平均値、分散具合と、Friedmanの分散分析手法による統計的有意差とを表している。この結果から、「楽しさ」、「キャラクタになった感覚」及び「キャラクタを操作している感覚」の3つの官能評価項目では、実施例1が比較例1より有意水準で1%高い評価を受けている。更に、「楽しさ」では、実施例1が比較例2より有意水準で1%高い評価を受けている。これに対し、「対話しやすさ」では、実施例1と比較例1又は比較例2との顕著な差は見られないが、アンケートによる回答では実施例1が比較例1より高い評価を受けており、比較例2に比べて有意水準5%で高い評価であった。

40

【0112】

次に、一対評価の結果について、図26に示す。図26は、例えば実施例1が比較例1又は比較例2に優る評価を得た場合、3列目の実施例1に対する1行目の比較例1、2行目の比

50

較例 2 の数値を加算した結果、すなわち実施例 1 , 比較例 1 及び比較例 2 の勝敗表を表している。この結果から、総合評価では実施例 1 比較例 1 比較例 2 の優劣関係にあり、比較例 1 及び比較例 2 に対して実施例 1 の評価が極めて高いことが窺える。

【 0 1 1 3 】

ここで、この一対評価の結果を定量的に評価するため、下記数 4 及び数 5 に示すBradley-Terryモデルを利用して、図26に見られるように、実施例 1、比較例 1 及び比較例 2 それぞれの強度 を算出している。この強度 の比較を表すグラフを図27に示す。ここで、数 4 中、 P_{ij} は*i*が*j*に優る確率、 π_i は*i*の強度を、 π_j は*j*の強度をそれぞれ表している。この度算出した強度 は、有意水準 5 %での適合度検定と尤度比検定とを実施してモデル整合性を検定した結果、モデル整合性は棄却されなかったので、妥当性が保証されている。これから、この強度 により実施例 1 , 比較例 1 及び比較例 2 を比べると、実施例 1、すなわち本発明が用いるリズム画像 5 の有効性が顕著であることが理解される。

【 0 1 1 4 】

【数 4】

$$P_{ij} = \frac{\pi_i}{\pi_i + \pi_j}$$

【 0 1 1 5 】

【数 5】

$$\sum \pi_i = const.(100)$$

【実施例 2】

【 0 1 1 6 】

次に、外部入力部としてデータグローブ614を用いた実施例 2 の有効性について、上記試験、すなわち 7 段階評価及び一対評価を実施した。外部入力部が携帯電話型キーボードからデータグローブ614に変わった以外、試験の装置構成や 7 段階評価及び一対評価の試験内容は上記実施例 1、比較例 1 及び比較例 2 の場合と同じである。会話時間は、7 段階評価については上記試験と同じ30秒ずつとしたが、一対評価ではデータグローブ614の操作性を考慮して、2分ずつとした。

【 0 1 1 7 】

実施例 2 は、話し手動作及び聞き手動作を混在させて動く本人アバター51の動画と話し手動作及び聞き手動作を混在させて動く相手アバター52の動画を重ね合わせたリズム画像 5 に、特別動作を追加している(図23参照)。また、比較例 3 は、実施例 2 のリズム画像における本人アバター及び相手アバターからそれぞれ特別動作を取り除いた画像で、前記本人アバター及び相手アバターは話し手動作及び聞き手動作からなるリズム動作だけを実行する。比較例 4 は、実施例 2 のリズム画像における本人アバター及び相手アバターからそれぞれリズム動作を取り除いた画像で、前記本人アバター及び相手アバターは特別動作だけを実行する。ここで、比較例 4 において、口の開閉動作や目の瞬き動作がなくなると不自然なので、口の開閉動作及び目の瞬き動作のみはリズムタイミングに基づいて実行させている。

【 0 1 1 8 】

本人アバター51及び相手アバター52の可動部及び前記各可動部の動きは、上述通りである。本人アバター51又は相手アバター52の特別動作は、各小型ノートパソコン 6 に接続したデータグローブ614(Immersion製 CyberGlove)を用いた。このデータグローブ614は、図 28に見られるように各モニターの右手に装着させ、親指が左前腕及び左上腕の上下の回転運動、人さし指及び中指が頭の前後の回転運動、人さし指及び中指の差が頭の左右の回転運動、薬指及び小指が右前腕及び右上腕の上下の回転運動、そして手首が胸(腰)の前後の回転運動を割り当てている。データグローブ614の各指に割り当てた左右の逆は、上述同様、画面に表示される本人アバター51及び相手アバター52を見たままに操作させるための便宜である。

【 0 1 1 9 】

データグローブ614を用いた場合の実施例2における7段階評価の結果を図29に、一対評価の結果を図30に、そして実施例2、比較例3及び比較例4それぞれの強度の比較を表すグラフを図31に示す。この結果から、7段階評価及び一対評価では上記試験同様の傾向が読み取れる。また、実施例2の強度は、実施例1の強度に比べて若干低い値を示しているが、これは操作が容易な携帯電話型キーボード613(図24参照)と、操作に慣れないデータグローブ614との操作性の相違に基づくものと見れば、実施例2は、実施例1と同等の有効性があると言える。

【図面の簡単な説明】

【0120】

【図1】本人アバターの動画と相手の実写動画とを重ね合わせたリズム画像を用いた携帯電話の使用状態を示す斜視図である。 10

【図2】本人アバターの動画と相手の実写動画とを重ね合わせたリズム画像を用いた携帯電話の本発明に係る構成を示すブロック図である。

【図3】リズム画像を生成する処理手順を表す処理フローチャートである。

【図4】本人アバターの話し手タイミングの推定と話し手動作の生成とを表す処理フローチャートである。

【図5】本人アバターの聞き手タイミングの推定と聞き手動作の生成とを表す処理フローチャートである。

【図6】本人アバターの特別タイミングの推定と特別動作の生成とを表す処理フローチャートである。 20

【図7】聞き手タイミングのマクロ推定とミクロ推定とを表すタイミングチャートである。

【図8】聞き手タイミングのマクロ推定とミクロ推定とを実行する論理回路のブロック図である。

【図9】本人アバターの正面図である。

【図10】話し手タイミング、聞き手タイミング及び特別タイミングからなるリズムタイミングを表すタイミングチャートである。

【図11】本人アバターの右前腕の基本的な動きを表す部分正面図である。

【図12】本人アバターの右前腕の動きの重なりを表す部分正面図である。

【図13】画面の表示例を表す本人が用いる携帯電話の部分正面図である。 30

【図14】特別動作を含むリズム動作に従って動く相手アバターの動画からなるリズム画像を用いた携帯電話の使用状態を示す斜視図である。

【図15】特別動作を含むリズム動作に従って動く相手アバターの動画からなるリズム画像を用いた携帯電話の本発明に係る構成を示すブロック図である。

【図16】リズム画像を生成する処理手順を表す処理フローチャートである。

【図17】画面の表示例を表す本人が用いる携帯電話の部分正面図である。

【図18】本例の携帯電話を応用した会話支援端末の本発明に係る構成を示すブロック図である。

【図19】本人アバター及び相手アバターの動画を組み合わせたリズム画像を用いた携帯電話の使用状態を示す斜視図である。 40

【図20】本人アバター及び相手アバターの動画を組み合わせたリズム画像を用いた携帯電話の本発明に係る構成を示すブロック図である。

【図21】リズム画像を生成する処理手順を表す処理フローチャートである。

【図22】画面の表示例を表す本人が用いる携帯電話の部分正面図である。

【図23】実施例1における試験風景を表す図である。

【図24】実施例1における本人アバター又は相手アバターの可動部と携帯電話型キーボードのテンキーとの対応関係を表す図である。

【図25】実施例1における7段階評価の結果を表すグラフである。

【図26】実施例1における一対評価の結果を表す表である。

【図27】実施例1における一対評価の結果を表すグラフである。 50

【図 28】実施例 2 における本人アバター又は相手アバターの可動部とデータグローブとの対応関係を表す図である。

【図 29】実施例 2 における 7 段階評価の結果を表すグラフである。

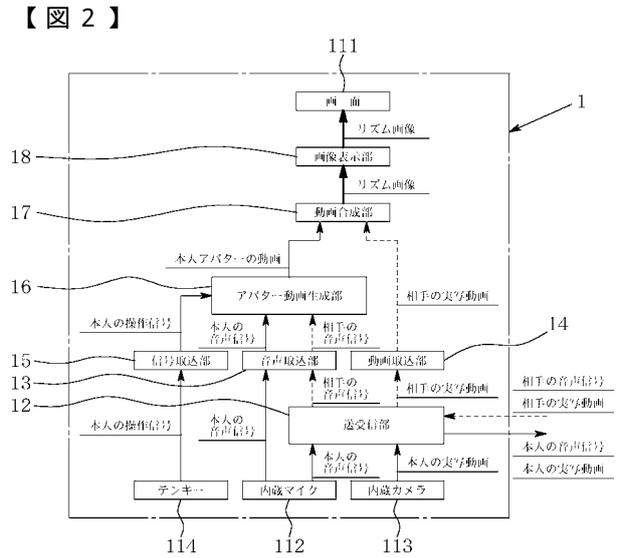
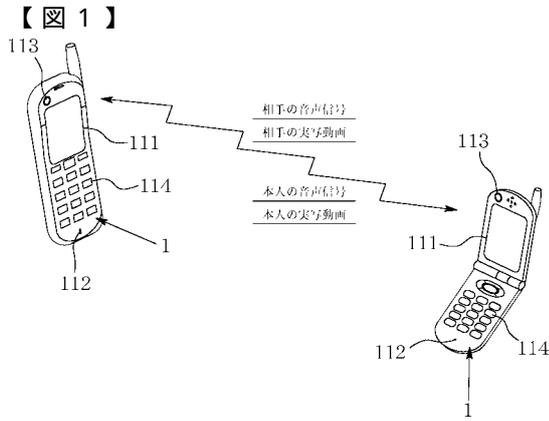
【図 30】実施例 2 における一対評価の結果を表す表である。

【図 31】実施例 2 における一対評価の結果を表すグラフである。

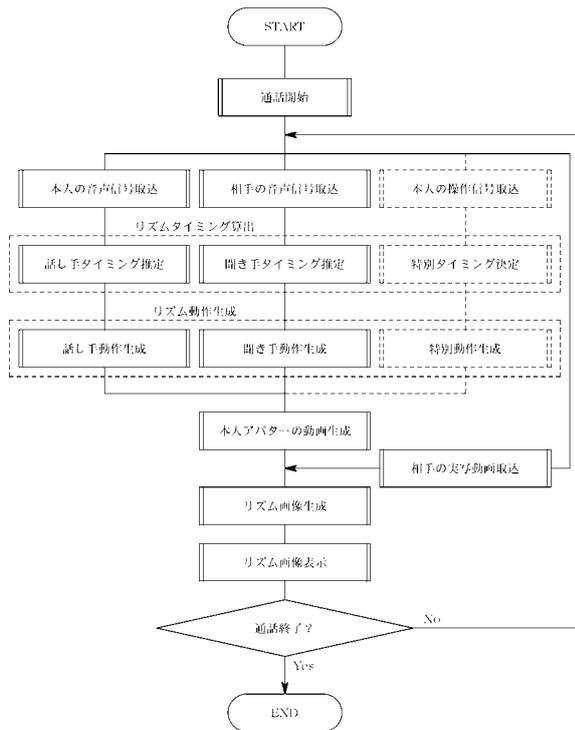
【符号の説明】

【 0 1 2 1 】

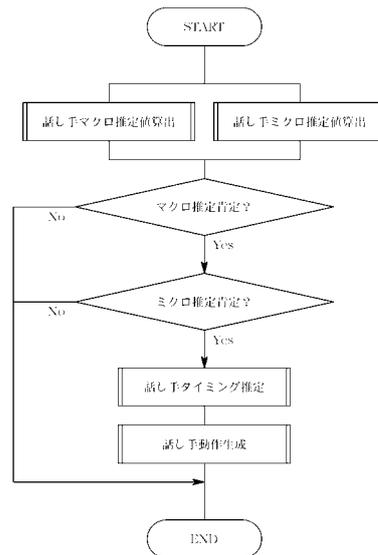
1	携帯電話	
111	画面	
112	内蔵マイク	10
113	内蔵カメラ	
114	テンキー	
12	送受信部	
13	音声取込部	
14	動画取込部	
15	信号取込部	
16	アバター動画生成部	
17	動画合成部	
18	画像表示部	
2	携帯電話	20
211	画面	
212	内蔵マイク	
214	テンキー	
22	送受信部	
23	音声取込部	
25	信号取込部	
26	アバター動画生成部	
28	画像表示部	
4	携帯電話	
411	画面	30
412	内蔵マイク	
414	テンキー	
42	送受信部	
43	音声取込部	
45	信号取込部	
46	アバター動画生成部	
47	動画合成部	
48	画像表示部	
5	リズム画像	
51	本人アバター	40
52	相手アバター	
53	相手（実写動画）	



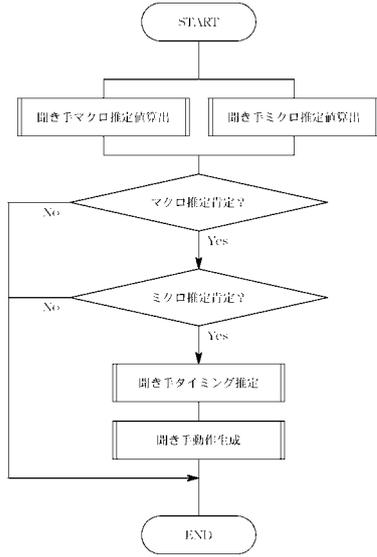
【図3】



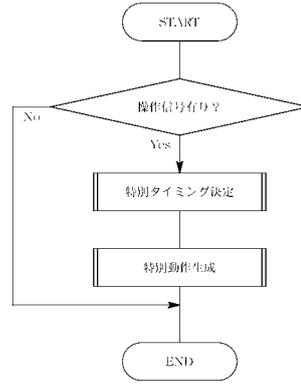
【図4】



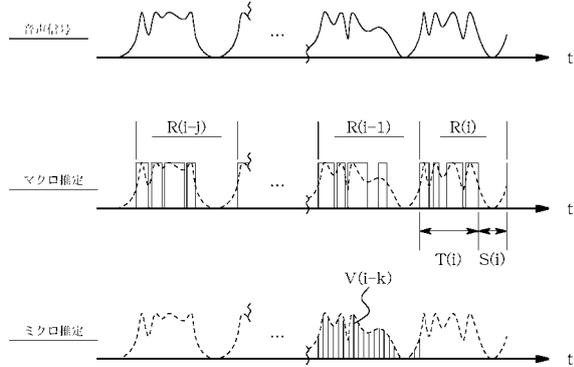
【図5】



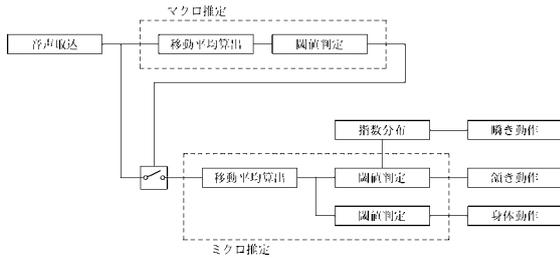
【図6】



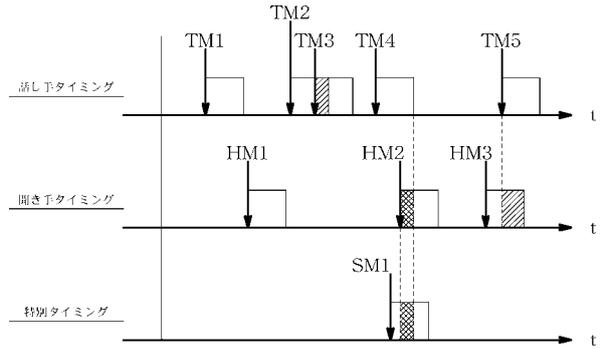
【図7】



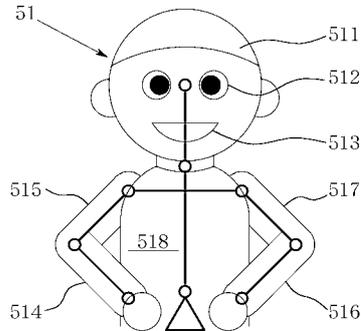
【図8】



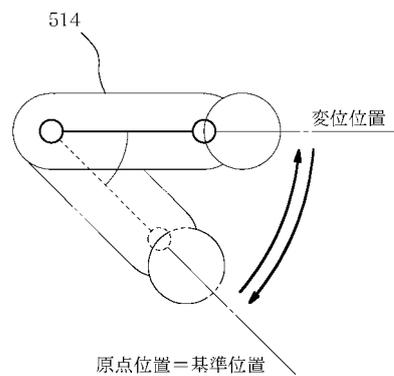
【図10】



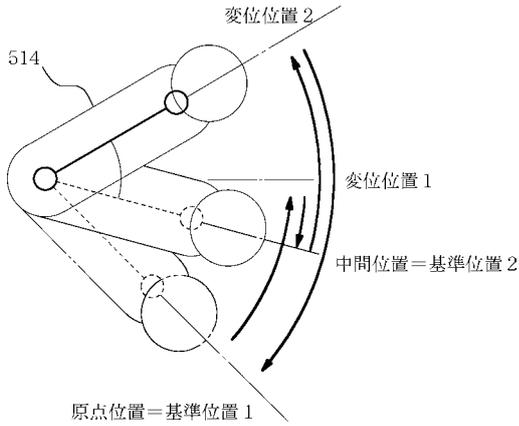
【図9】



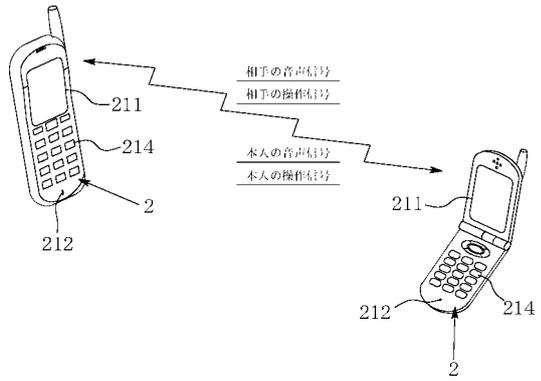
【図11】



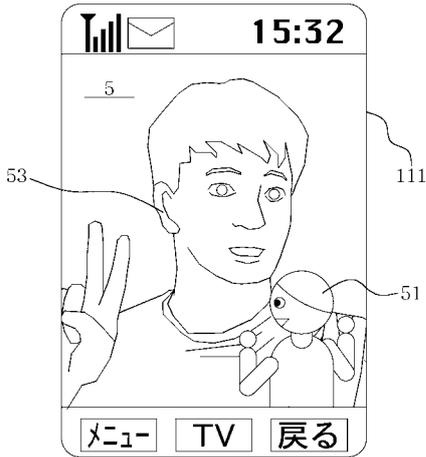
【図12】



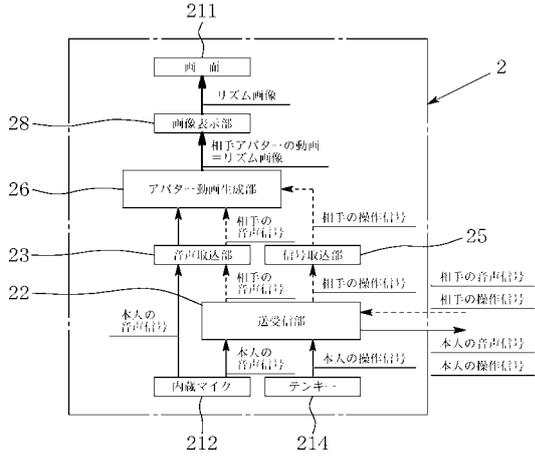
【図14】



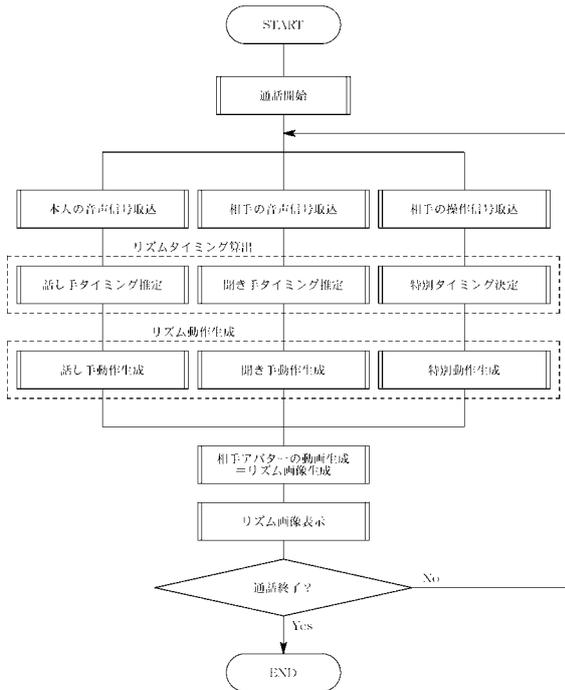
【図13】



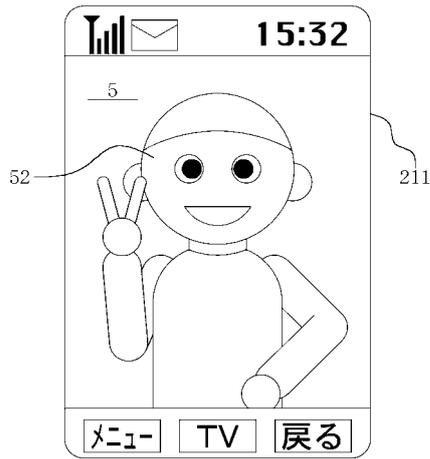
【図15】



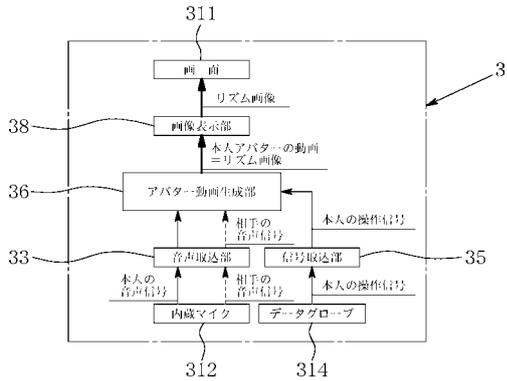
【図16】



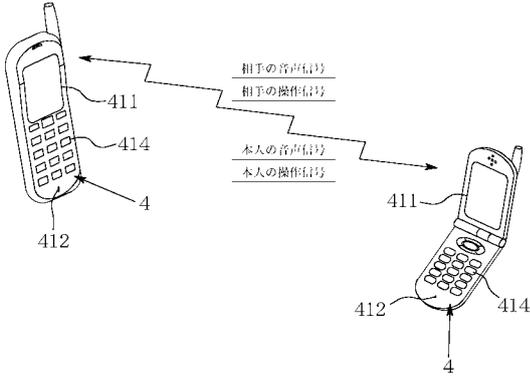
【図17】



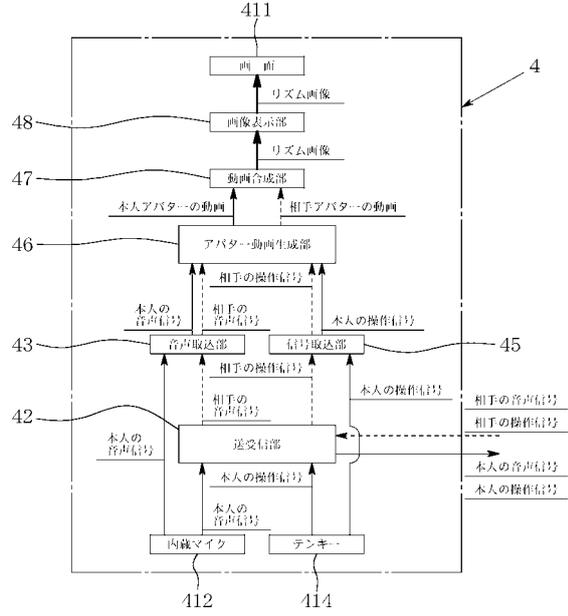
【図18】



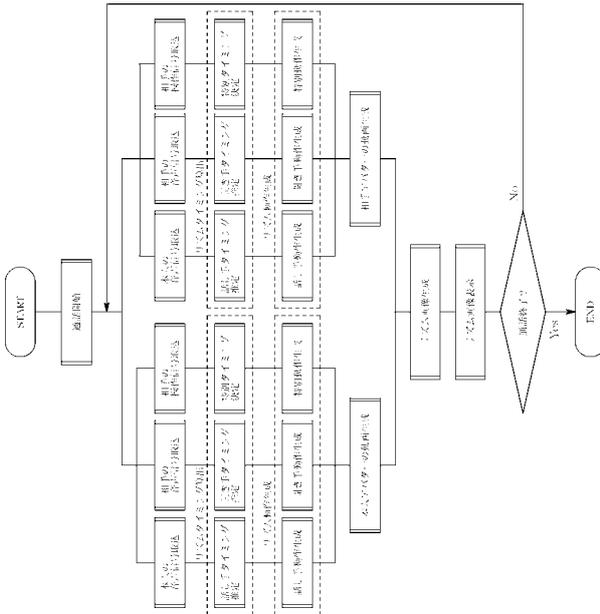
【図 19】



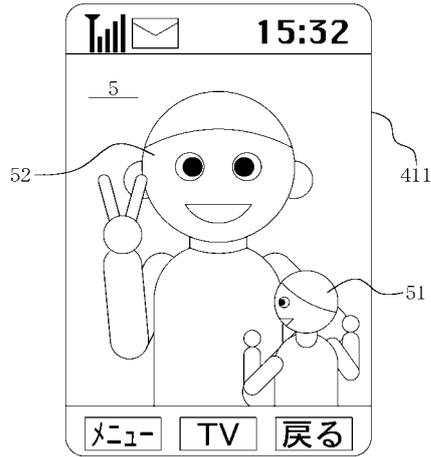
【図 20】



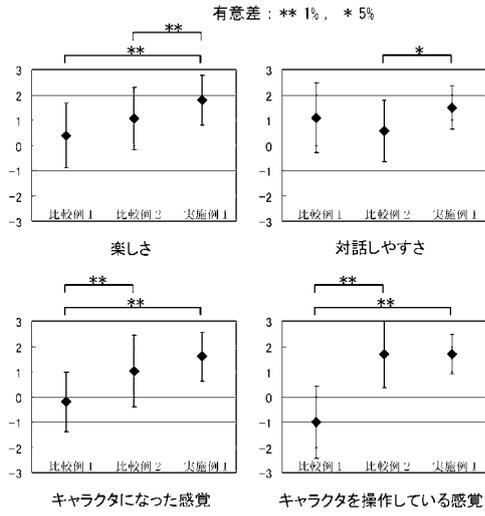
【図 21】



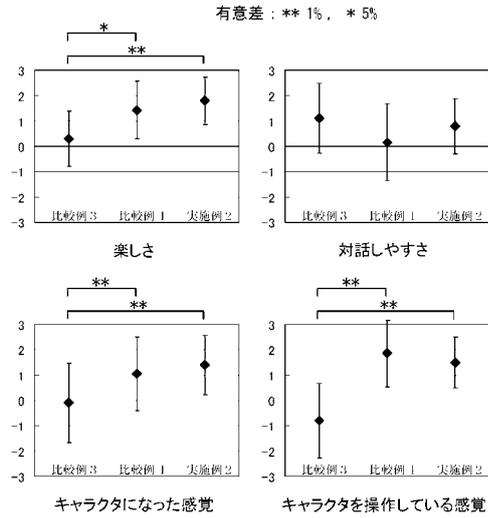
【図 22】



【図25】



【図29】



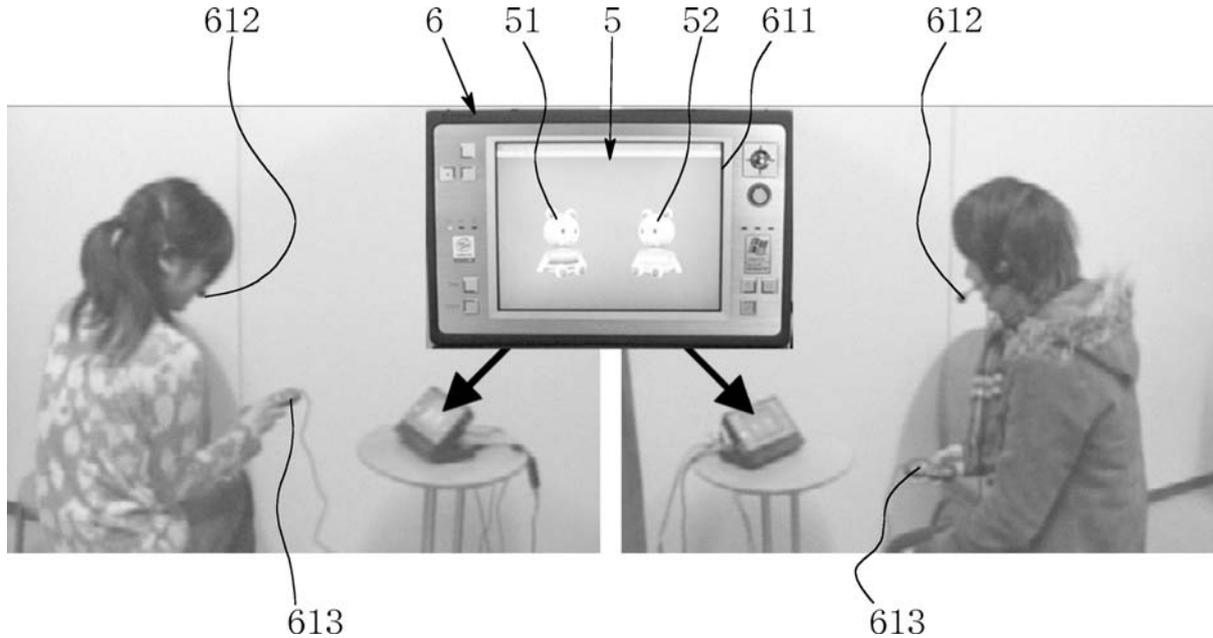
【図26】

	比較例1	比較例2	実施例1	π
比較例1		16	4	6.88
比較例2	24		3	9.32
実施例1	36	37		83.8

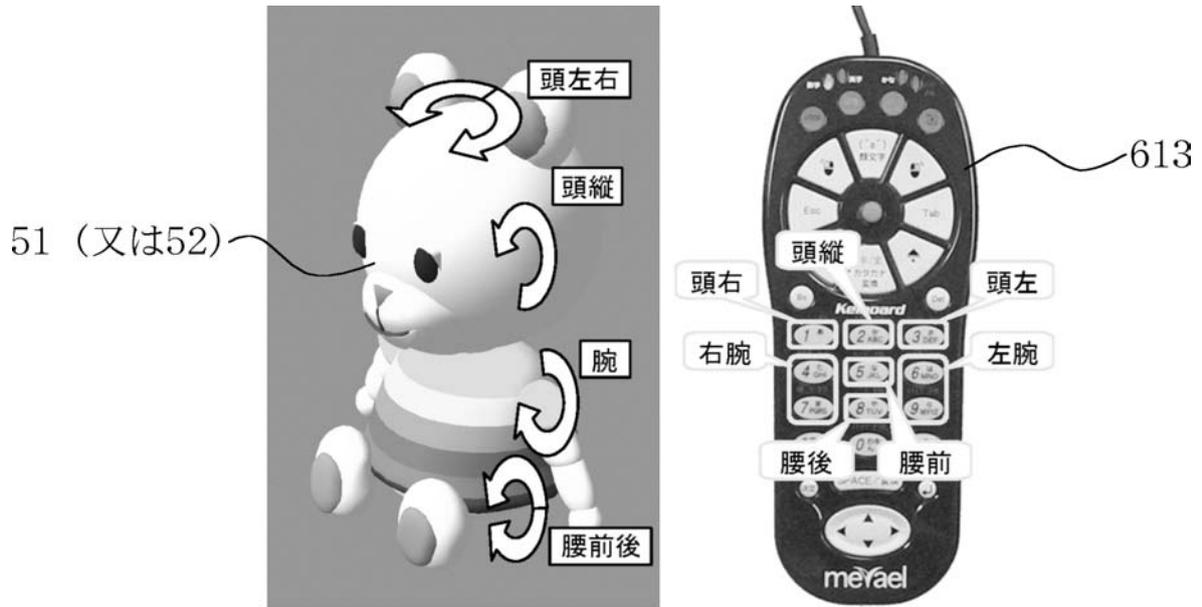
【図30】

	比較例3	比較例4	実施例2	π
比較例3		18	4	8.34
比較例4	22		4	9.89
実施例2	36	36		81.7

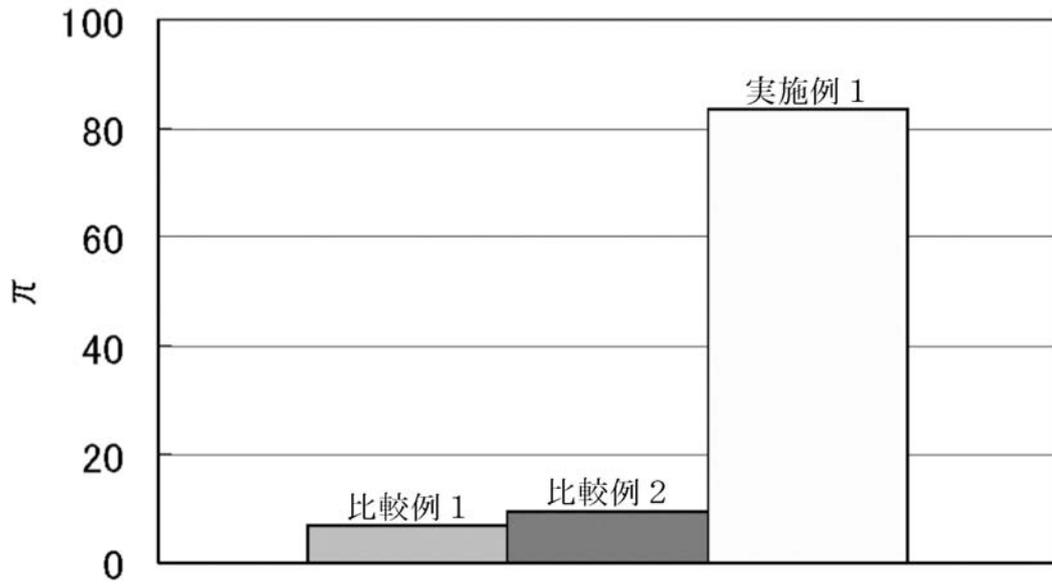
【図23】



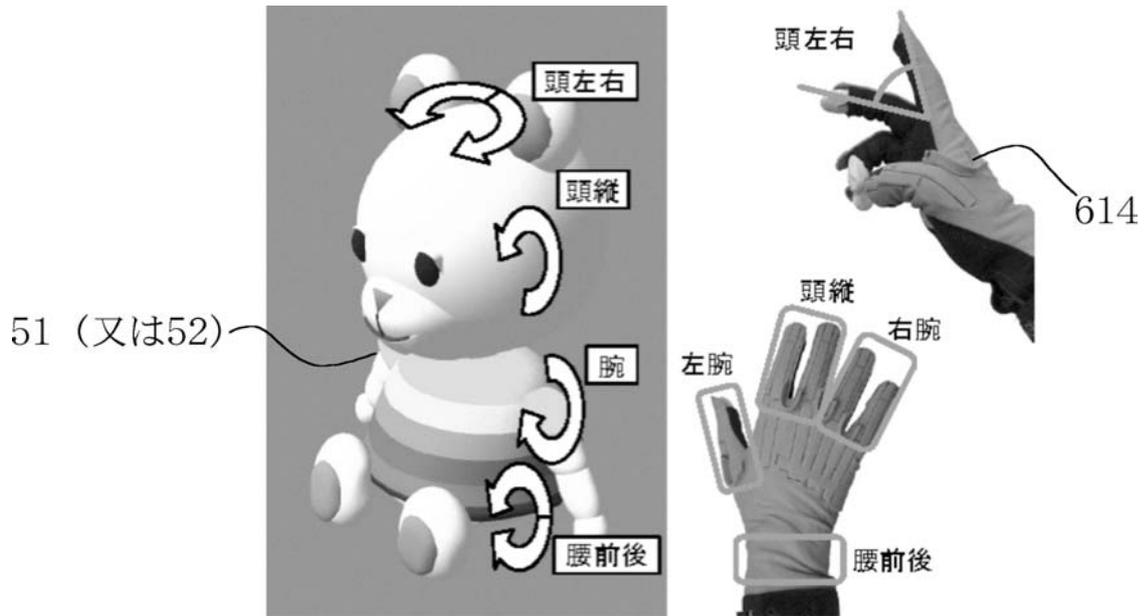
【図24】



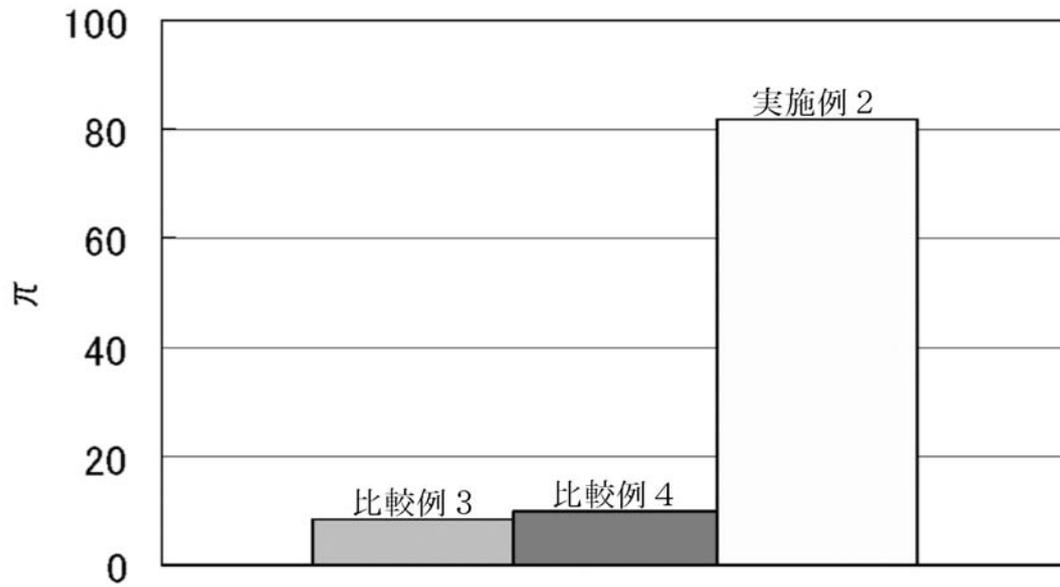
【図27】



【図28】



【図31】



フロントページの続き

(72)発明者 大崎 浩司
岡山県総社市窪木 1 1 1 番地 岡山県立大学内

審査官 真木 健彦

(56)参考文献 特開 2 0 0 3 - 1 0 8 5 0 2 (J P , A)
特開平 0 7 - 2 5 5 0 4 4 (J P , A)
特開 2 0 0 1 - 1 5 7 1 8 4 (J P , A)

(58)調査した分野(Int.Cl. , DB名)
G 0 6 T 1 3 / 4 0
G 0 6 T 1 3 / 8 0
G 0 6 T 1 9 / 0 0