

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第3598384号

(P3598384)

(45) 発行日 平成16年12月8日(2004.12.8)

(24) 登録日 平成16年9月24日(2004.9.24)

(51) Int. Cl.⁷

F I

G 1 1 B 20/00

G 1 1 B 20/00 G

G 1 1 B 20/10

G 1 1 B 20/00 Z

G 1 1 B 31/00

G 1 1 B 20/10 3 1 1

H O 4 N 7/15

G 1 1 B 31/00 5 1 7

H O 4 R 3/00

H O 4 N 7/15

請求項の数 13 (全 12 頁) 最終頁に続く

(21) 出願番号 特願2002-269608 (P2002-269608)
 (22) 出願日 平成14年9月17日(2002.9.17)
 (65) 公開番号 特開2004-110898 (P2004-110898A)
 (43) 公開日 平成16年4月8日(2004.4.8)
 審査請求日 平成14年9月17日(2002.9.17)

特許法第30条第1項適用 平成14年3月18日-20日において開催された言語処理学会第8回年次大会において発表

(73) 特許権者 301022471
 独立行政法人情報通信研究機構
 東京都小金井市貫井北町4-2-1
 (74) 代理人 100085338
 弁理士 赤澤 一博
 (74) 代理人 100118245
 弁理士 井上 敬子
 (72) 発明者 高梨 克也
 東京都小金井市貫井北町4-2-1 独立
 行政法人通信総合研究所内
 (72) 発明者 井佐原 均
 東京都小金井市貫井北町4-2-1 独立
 行政法人通信総合研究所内

最終頁に続く

(54) 【発明の名称】 三者会話データ収録装置

(57) 【特許請求の範囲】

【請求項1】

三者間における会話を収録する三者会話データ収録装置であって、各話者の音声を各々収録する収録部と、前記収録部で収録する各話者の音声を異なる二者毎の組にして録音する録音部と、前記収録部で収録する他の二者の音声をその収録時に一の話者各々に対して出力する音声出力部とを具備することを特徴とする三者会話データ収録装置。

【請求項2】

前記収録部が、各々の話者の音声を他の話者の音声と混同しないように収録することを特徴とする請求項1記載の三者会話データ収録装置。

【請求項3】

各話者毎に対をなす二つの音声出力部を備え、この対をなす二つの音声出力部を、各話者の両耳にそれぞれ直接取り付けあるいは近傍に配置していることを特徴とする請求項1又は2記載の三者会話データ収録装置。

【請求項4】

前記音声出力部が、前記収録部で収録する他の話者の音声出力レベルを、前記他の話者が位置する方向の耳に対して大きく前記他の話者が位置しない方向の耳に対して小さくなるようにレベル調整し一の話者の両耳に対して出力することを特徴とする請求項3記載の三者会話データ収録装置。

【請求項5】

他の話者を映した映像を表示する画像出力部を備え、前記音声出力部が、前記収録部で収

10

20

音する他の話者の音声出力レベルを、前記画像出力部に表示する他の映像の位置に合致させて出力することを特徴とする請求項4記載の三者会話データ収録装置。

【請求項6】

他の話者を映した映像を表示する画像出力部を備え、前記音声出力部が、前記収録部で収録する他の二者の音声を、前記画像出力部に表示される前記他の二者の話者間における左右の位置関係と一致させて一の話者の両耳に出力することを特徴とする請求項3記載の三者会話データ収録装置。

【請求項7】

前記音声出力部を、自分の音声が自分に聞こえるように、前記収録部で収録する話者の音声をその話者へ出力するように構成していることを特徴とする請求項1、2、3、4、5又は6記載の三者会話データ収録装置。

10

【請求項8】

前記音声出力部が、三者間の会話を促す第四の話者の音声を前記各話者に出力することを特徴とする請求項1、2、3、4、5、6又は7記載の三者会話データ収録装置。

【請求項9】

前記録音部の録音レベルや前記音声出力部の音声出力レベルを調整可能な正弦波等の基準信号を出力する基準信号出力部を具備することを特徴とする請求項1、2、3、4、5、6、7又は8記載の三者会話データ収録装置。

【請求項10】

会話時の前記各話者の音声と各話者が映った映像とを収録する総合収録部を備えていることを特徴とする請求項1、2、3、4、5、6、7、8又は9記載の三者会話データ収録装置。

20

【請求項11】

前記録音部で収録する音声が、所定のサンプリング周波数でサンプリングし所定の値で量子化してデジタルデータ化したものであることを特徴とする請求項1、2、3、4、5、6、7、8、9又は10記載の三者会話データ収録装置。

【請求項12】

前記録音部で収録する音声が、記録媒体における録音位置を絶対的あるいは相対的に指定可能なタイムコードとともに収録したものであることを特徴とする請求項11記載の三者会話データ収録装置。

30

【請求項13】

前記録音部が、異なる二者の音声を各チャンネルに分離して収録する3台の2チャンネル録音機で構成したものであることを特徴とする請求項12記載の三者会話データ収録装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、三者間における対話進行のメカニズムの解明を行う際に用いる三者会話データ収録装置に関するものである。

【0002】

40

【従来の技術】

近年、インターネットがビジネスと同様、生活にも使われ始められるようになり、今後、健常者、高齢者、障害者を含め益々多くの人々が利用・参加できるものにするために、人間とコンピュータとの間における会話を行う対話システムの開発が活発に行われるようになってきている。このような対話システムの開発には、対話進行のメカニズムの解明に必要な不可欠な対話コーパスが用いられている（例えば、特許文献1参照。）。

【0003】

【特許文献1】

特表2001-516905号公報（段落番号0002～0004）

【0004】

50

【発明が解決しようとする課題】

しかしながら、従来の対話コーパスは二者間の課題指向的対話を収録したものが大半であり、対話の基本構造を定式化する上で、二者対話に基づく分析では不十分な点が多く、三者以上の参加者を含むコミュニケーションツールの開発や対話規則についての理論的研究のための基礎的データとして、三者会話コーパスの構築が重要とされているが、この三者間における会話を収録するに有用な収録装置の提供がなされていなかった。

【0005】

そこで、本発明は、上述する問題を解決することを主たる課題とするものである。

【0006】**【課題を解決するための手段】**

すなわち、本発明は、三者間における会話を収録する三者会話データ収録装置であって、各話者の音声を各々収録する収録部と、前記収録部で収録する各話者の音声を異なる二者毎の組にして録音する録音部と、前記収録部で収録する他の二者の音声をその収録時に一の話者各々に対して出力する音声出力部とを具備することを特徴とする。

【0007】

このような構成によれば、録音部に組みにして録音した二者毎の組にした話者間における対話進行を分析でき、さらに、この分析した二者毎の話者間における対話進行に基づき、三者間における対話進行の過程の分析を好適に行うことができるといった、対話コーパスの作成に有用な三者会話データ収録装置を提供することができる。

【0008】

なお、複数の話者が重複して発話した場合でも、収録した各話者の音声の聞き分けを確実に行うためには、前記収録部が、各々の話者の音声を他の話者の音声と混同しないように収録することが望ましい。

【0009】

また、前記収録部が各話者の収録時に、音響出力部から出力される他者の音声の混同することを防止するためには、各話者毎に対をなす二つの音声出力部を備え、この対をなす二つの音声出力部を、各話者の両耳にそれぞれ直接取り付けあるいは近傍に配置していることが好ましい。

【0010】

さらに、前記音声出力部の取り付け状態による不自然さ等を感じさせず自然な会話を促すためには、前記音声出力部が、前記収録部で収録する他の話者の音声出力レベルを、前記他の話者が位置する方向の耳に対して大きく前記他の話者が位置しない方向の耳に対して小さくなるようにレベル調整し一の話者の両耳に対して出力し、さらに、他の話者を映した映像を表示する画像出力部を備え、前記音声出力部が、前記収録部で収録する他の話者の音声出力レベルを、前記画像出力部に表示する他者の映像の位置に合致させて出力すればよい。また、他の話者を映した映像を表示する画像出力部を備え、前記音声出力部が、前記収録部で収録する他の二者の音声を、前記画像出力部に表示される前記他の二者の話者間における左右の位置関係と一致させて一の話者の両耳に出力するといった態様も考えられる。

【0011】

なお、話者に違和感等を覚えさせることなく自然な会話を行わせるためには、前記音声出力部を、自分の音声が自分に聞こえるように、前記収録部で収録する話者の音声をその話者へ出力するように構成すればよく、さらに、三者間における円滑な会話を実現させるためには、前記音声出力部が、三者間の会話を促す第四の話者の音声を前記各話者に出力することが望まれる。

【0012】

また、本発明の三者会話データ収録装置を構成する各部のレベル調整を行うためには、前記録音部の録音レベルや前記音声出力部の音声出力レベルを調整可能な正弦波等の基準信号を出力する基準信号出力部を具備すればよい。

【0013】

10

20

30

40

50

さらに、会話の全体を好適に把握するためには、会話時の前記各話者の音声と各話者が映った映像とを収録する総合収録部を備えればよい。

【0014】

なお、収録した音声データを好適に整理・分析するためには、前記録音部で収録する音声、所定のサンプリング周波数でサンプリングし所定の値で量子化してデジタルデータ化したものであることが好ましく、特に、前記録音部で収録する音声が、記録媒体における録音位置を絶対的あるいは相対的に指定可能なタイムコードとともに収録したものであることが望まれる。

【0015】

また、前記録音部に、異なる二者の音声を各チャンネルに分離して収録する3台の2チャンネル録音機を用いれば、本発明の三者会話データ収録装置を簡単に構成することができる。

10

【0016】

【発明の実施の形態】

以下、図面を用いて本発明の実施形態について説明する。

【0017】

図1は、本発明の三者会話データ収録装置Aを構成する音声系統図を示す図である。図2は、本発明の三者会話データ収録装置Aを構成する映像系統図を示す図である。図3は、本発明の三者会話データ収録装置Aにおける収録の実施態様を示す図である。

【0018】

20

本発明の三者会話データ収録装置Aは、図1、図2及び図3に示すように、3人の話者1a、1b、1c（以下、「話者1」と総称する。）それぞれの音声A、B、Cを収録する収録部たるマイクロホン2a、2b、2c（以下、「マイクロホン2」と総称する。）と、このマイクロホン2で収録する各話者1の音声A、B、Cを異なる二者毎の組みにして録音する録音部3と、各話者1の両耳に取り付ける対をなす音声出力部たる音声出力ユニットを設けたヘッドホン4a、4b、4c（以下、「ヘッドホン4」と総称する。）と、前記話者1の会話を促す第4の話者たるオペレータ1dの音声Dを収録するマイクロホン2dと、前記マイクロホン2で収録する話者1の音声Dを前記オペレータ1dに対して出力するヘッドホン4dと、前記収録部2で収録する各話者1の音声A、B、Cを調整し前記録音部3やヘッドホン4に出力する音響調整部5と、前記録音部3やヘッドホン4のレベル調整を行うための基準信号を出力する基準信号出力部6と、前記話者1の音声等を収録し会話全体の全体像を把握するための総合収録部7とを主な構成要素としている。そして、各々の話者1の音声A、B、Cが他の話者1の音声A、B、Cと混同しないように各話者1をそれぞれ入室させる3室から成る防音ボックスSa、Sb、Sc（以下、「防音ボックスS」と総称する。）を設け、前記各部の全体又はその主要部をこの防音ボックスS内に配置するとともに、各話者1の対面する環境を創出するための、各話者1a、1b、1cをそれぞれ撮影するビデオカメラVa、Vb、Vc（以下、「ビデオカメラV」と総称する。）と、このビデオカメラVで映した他者の映像を表示する画像出力部8とを備えている。

30

【0019】

40

以下、各部を詳述する。

【0020】

マイクロホン2は、各話者1毎の音声A、B、Cを収録するものであって、話者1aの音声Aを収録するマイクロホン2aと、話者1bの音声Bを収録するマイクロホン2bと、話者1cの音声Cを収録するマイクロホン2cとを、それぞれの話者1a、1b、1cの近傍に配置している。

【0021】

録音部3は、異なる二者の音声を各チャンネルに分離して収録する入力チャンネルを2チャンネル有する3台の2チャンネル録音機3a、3b、3cで構成されるものである。なお、本実施形態では、この2チャンネル録音機に一般的なDATレコーダを用い、このD

50

A Tレコーダに挿入する記録媒体たるデジタルオーディオテープ(以下、「DAT」とする。)に各音声を収録するようにしている。より具体的には、2チャンネル録音機3 aには、一方のチャンネルに話者1 aの音声Aを他方のチャンネルに話者1 bの音声Bを録音するように設定している。2チャンネル録音機3 bには、一方のチャンネルに話者1 bの音声Bを他方のチャンネルに話者1 cの音声Cを録音するように設定している。2チャンネル録音機3 cには、一方のチャンネルに話者1 cの音声Cを他方のチャンネルに話者1 aの音声Aを録音するように設定している。なお、本実施形態では、前記2チャンネル録音機3 a、3 b、3 cの各チャンネルにおいて、各話者1の音声をサンプリング周波数が48 kHz、16 bitで量子化したデジタルデータでDATに録音するように設定している。

10

【0022】

ヘッドホン4は、前記マイクロホン2で収音する他の二者の音声をその収音時に一の話者1各々に対して出力するものであって、各話者1の両耳に取り付ける各々のヘッドホン4には、それぞれ対をなし異なる出力レベルで出力可能な音声出力ユニットが設けられている。より具体的には、話者1 aの左耳及び右耳にそれぞれ音声出力ユニット4 a L、4 a Rが取り付けられ、話者1 bの左耳及び右耳にそれぞれ音声出力ユニット4 b L、4 b Rが取り付けられ、話者1 cの左耳及び右耳にそれぞれ音声出力ユニット4 c L、4 c Rが取り付けられるように構成している。また、本実施形態では、オペレータ1 dの左耳及び右耳にそれぞれ同様の音声出力ユニット4 d L、4 d Rを設けている。

【0023】

音響調整部5は、複数の音声や映像を混合したり調整したりする一般的な音響ミキサを一又は複数台接続することにより、前記マイクロホン2で収音する各話者1の音声A、B、Cを混合したりその周波数特性等を調整したりするものであって、本実施形態では、図示しない五台の音響ミキサのうち、四台を各話者1 a、1 b、1 c及びオペレータ1 dの近傍に配置し、他の一台を前記四台のミキサにそれぞれ接続し得るように配置して、これら五台の音響ミキサを適直接続することにより、この音響調整部5を構成している。そして、この音響調整部5に、前記マイクロホン2で収音する各話者1の音声A、B、C及びマイクロホン2 dで収音するオペレータ1 dの音声Dと前記基準信号出力部6が出力する基準信号等を入力する入力チャンネルと、この入力チャンネルに入力される前記音声A、B、C、D等を混合したりその周波数特性等を調整したりする図示しない音響特性調整部と、この音響特性調整部で調整等が行われた出力信号を出力する出力チャンネルとを備えるようにしている。より具体的には、前記音響特性調整部を適宜調整することにより、入力チャンネルから入力する話者1 aの音声Aを、2チャンネル録音機3 aの一方のチャンネル及び2チャンネル録音機3 cの一方のチャンネルに対して、他の話者1 b、1 cの音声B、Cと分離した状態で出力するように設定している。同様に、入力チャンネルから入力する話者1 bの音声Bを、2チャンネル録音機3 aの一方のチャンネル及び2チャンネル録音機3 bの一方のチャンネルに対して、他の話者1 a、1 cの音声A、Cと分離した状態で出力するように設定している。また、入力チャンネルから入力する話者1 cの音声Cを、2チャンネル録音機3 bの一方のチャンネル及び2チャンネル録音機3 cの一方のチャンネルに対して、他の話者1 a、1 bの音声A、Bと分離した状態で出力するように設定している。

20

30

40

【0024】

さらに、前記音響特性調整部を適宜調整することにより、各話者の両耳に直接取り付けられている対をなす二つの音声出力ユニットそれぞれに、画像出力部8に表示する他の二者の映像と一致するように出力レベルを調節して出力するように設定している。より具体的には、本実施形態では、後述する画像出力部8に表示する他の二者が映った映像と一致させ、話者1 aの左耳に直接取り付けた音声出力ユニット4 a Lには話者1 cの音声Cを出力し、その右耳に直接取り付けた音声出力ユニット4 a Rには話者1 bの音声Bを出力するように設定している。同様に、話者1 bの左耳に直接取り付けた音声出力ユニット4 b Lには話者1 aの音声Aを出力し、その右耳に直接取り付けた音声出力ユニット4 b Rには話

50

者 1 c の音声 C を出力するように設定している。また、話者 1 c の左耳に直接取り付けられた音声出力ユニット 4 c L には話者 1 b の音声 B を出力し、その右耳に直接取り付けられた音声出力ユニット 4 c R には話者 1 a の音声 A を出力するように設定している。

【 0 0 2 5 】

さらにまた、本実施形態では、前記音響特性調整部を適宜調整することにより、各話者 1 が自分の音声を自分でも聴くことができるように各自のヘッドホン 4 に各自の音声を出力するように設定するとともに、オペレータ 1 d の音声 D を、各話者 1 のヘッドホン 4 及び該オペレータのヘッドホン 4 d に出力するように設定している。

【 0 0 2 6 】

基準信号出力部 6 は、前記録音部 3 の録音レベルや前記音声出力ユニットの音声出力レベルを調整可能な正弦波等の基準信号を出力するものである。また、本実施形態では、該基準信号出力部 6 が、収録する音声の D A T における録音位置を絶対的あるいは相対的な収録時間として示すタイムコードを収録時に出力し、このタイムコードが、前記録音部 3 で音声とともに D A T に収録されるように構成している。

10

【 0 0 2 7 】

総合収録部 7 は、会話時の前記各話者 1 の音声を 2 チャンネルの音声チャンネルに収録するとともに各話者 1 が映った映像を 1 チャンネルの映像チャンネルに収録するものである。なお、本実施形態では、各話者 1 の音声及び映像により会話全体の全体像を把握し得るように、この総合収録部 7 を一般的なデジタルビデオカメラ D V で構成し、そして一本のデジタルビデオカメラ用テープにおける一方の音声チャンネルに話者 1 a、1 c の音声 A 及び C を収録し、他方の音声チャンネルに話者 1 b とオペレータ 1 d の音声 B 及び D を収録するとともに、映像チャンネルにモニタ画面を 4 分割したそれぞれの画面に話者 1 a、1 b、1 c の映像が表示し得るように収録している。

20

【 0 0 2 8 】

防音ボックス S は、その内部で発生する音を外部に放射しないようにすると共に、その外部で発生した音を内部に進入しないようにする遮音機能を有するものであって、本実施形態では、防音ボックス S a、S b、S c にそれぞれ入出した話者 1 a、1 b、1 c の音声が、前記収録部 2 や音響出力部以外によっては相互に伝わらないように構成している。

【 0 0 2 9 】

ビデオカメラ V は、動画などを撮影できないいわゆる一般的なビデオカメラであって、本実施形態では、前記防音ボックス S a、S b、S c にそれぞれビデオカメラ V a、V b、V c を配置することにより、その防音ボックス S 内にいる話者 1 a、1 b、1 c それぞれの映像を撮影するように構成している。

30

【 0 0 3 0 】

画像出力部 8 は、前記ビデオカメラ V で撮影した話者 1 の映像を表示できないいわゆる一般的なモニタ装置であって、本実施形態では、前記防音ボックス S a、S b、S c にそれぞれ 2 台の画像出力部 8 を配置するとともに前記オペレータ 1 d が視ることが可能な位置に 1 台の画像出力部 8 x を配置している。より具体的には、話者 1 a が入室する防音ボックス S a には 2 台の画像出力部 8 a L と 8 a R とをそれぞれ話者 1 a に対して左右に並べて配置し、さらに、この画像出力部 8 a L には話者 1 c の映像 c を表示し画像出力部 8 a R には話者 1 b の映像 b を表示するように設定している。また、同様に、話者 1 b が入室する防音ボックス S b には 2 台の画像出力部 8 b L と 8 b R とをそれぞれ話者 1 b に対して左右に並べて配置し、さらに、この画像出力部 8 b L には話者 1 a の映像 a を表示し画像出力部 8 b R には話者 1 c の映像 c を表示するように設定し、話者 1 c が入室する防音ボックス S c には 2 台の画像出力部 8 c L と 8 c R とをそれぞれ話者 1 c に対して左右に並べて配置し、さらに、この画像出力部 8 c L には話者 1 b の映像 b を表示し画像出力部 8 c R には話者 1 a の映像 a を表示するように設定している。一方、前記オペレータ 1 d が視聴可能な位置に配置した画像出力部 8 x には、デジタルビデオカメラ D V から出力される各話者 1 の映像 a、b、c をモニタ画面上に 4 分割して表示するように設定している。

40

【 0 0 3 1 】

50

次に、以上のように構成される三者会話データ収録装置 A における収録方法について説明する。

【0032】

まず、基準信号出力部 6 から所定の周波数の正弦波を出力し、音響調整部 5 における図示しない音響特性調整部を調整することにより、前記録音部 3 の録音レベルや音声出力ユニットの音声出力レベルを調整する。なお、本実施形態では、2チャンネル録音機 3 a、3 b、3 c における各チャンネルの録音レベルが略等しくなるように設定するとともに、対をなす音声出力ユニット間における音声出力レベルが略等しくなるように設定している。

【0033】

そして、このように各レベルを調整した三者会話データ収録装置 A において、各話者 1 が初対面の組み合わせの場合にも、自然な活発な会話が行われるようオペレータ 1 d を用意し、その際の各話者 1 の音声や話者の身振りや表情を収録する。なお、オペレータ 1 d を用意せず、趣味や好み等各話者が話題にしやすい、自己紹介的な話題リストや写真等を用意することにより、自然な活発な会話が行われるように構成する実施態様も考えられる。

【0034】

この収録の具体的な方法としては、例えば、話者 1 a が他の話者 1 b、1 c に対して話し掛けると、話し掛けた音声 A は話者 1 a の近傍に設けたマイクロホン 2 a で收音され、このマイクロホン 2 a と接続している音響調整部 5 の入力チャンネルに入力される。そして、この音響調整部 5 に入力された話者 1 a の音声 A は、音響調整部 5 の音響特性調整部を介して、話者 1 b の左耳に取り付けている音声出力ユニット 4 b L 及び話者 1 c の右耳に取り付けている音声出力ユニット 4 c R に出力される。このようにこれら音声出力ユニット 4 b L 及び 4 c R から出力される話者 1 a の音声 A を聞いた話者 1 b 及び 1 c が、前記話者 1 a の音声 A の内容に応じた返事等を行うと、話者 1 b 及び 1 c が発する音声 B 及び C は、話者 1 a の音声と同様、音響調整部 5 の音響特性調整部を介して、他の話者に出力される。このようにして、三者間における会話が進行されることとなる。ところで、このような三者間における会話が進行している際に、話者 1 a の音声 A は、前記音響調整部 5 の音響特性調整部を介して 2チャンネル録音機 3 a 及び 3 c に出力され、それぞれの 2チャンネル録音機の 1チャンネルに、基準信号出力部 6 が出力するタイムコードとともに録音されることとなる。また、同様に、話者 1 b の音声 B は、2チャンネル録音機 3 a 及び 3 b に出力され、それぞれの 2チャンネル録音機の 1チャンネルに録音され、話者 1 c の音声 B は、2チャンネル録音機 3 b 及び、3 c に出力され、それぞれの 2チャンネル録音機の 1チャンネルに基準信号出力部 6 が出力するタイムコードとともに録音されることとなる。このようにして、三者間における会話の進行中に、各話者 1 の音声 A、B、C を録音することができる。

【0035】

ところで、このような会話の際に、各話者 1 の映像をその各話者 1 が入室している各防音ボックス S 内に設けたビデオカメラ V が撮影し、各防音ボックス S 内に設けた画像出力部 8 に表示する。より具体的には、ビデオカメラ V a で撮影する話者 1 a の映像は、話者 1 b の入室している防音ボックス S b 内の左側の画像出力部 8 b L 及び話者 1 c の入室している防音ボックス S b 内の右側の画像出力部 8 c R に出力される。同様に、ビデオカメラ V b で撮影する話者 1 b の映像は、話者 1 a の入室している防音ボックス S a 内の右側の画像出力部 8 a R 及び話者 1 c の入室している防音ボックス S c 内の左側の画像出力部 8 c L に出力され、ビデオカメラ V c で撮影する話者 1 c の映像は、話者 1 a の入室している防音ボックス S a 内の左側の画像出力部 8 c L 及び話者 1 b の入室している防音ボックス S b 内の右側の画像出力部 8 b R に出力される。

【0036】

このように、各防音ボックス S 内において画像出力部 8 に表示される他の話者の映像の方向と音声出力ユニットに出力される音声の方向とが一致しているため、不自然さ等を感じさせず自然な会話を促されることとなる。なお、各話者 1 の両耳に取り付けた音声出力ユニットの音声出力レベルを、前記他の話者が位置する方向の耳に対して大きく前記他の話

10

20

30

40

50

者が位置しない方向の耳に対して小さくなるようにレベル調整し一の話者の両耳に対して出力すればさらに自然な会話を促すこともできる。より具体的には、例えば、話者1 aの両耳に取り付けた音声出力ユニット4 a L及び4 a Rにおいて、話者1 cの音声Cを、この音声出力ユニット4 a Lで大きく音声出力ユニット4 a Rで小さくなるように出力し、同様に話者1 bの音声Bを、この音声出力ユニット4 a Lで小さく音声出力ユニット4 a Rで大きくなるように出力するようによい。

【0037】

以上に詳述したように、本実施形態の三者会話データ収録装置Aは、録音部3に組みにして録音した二者毎の組にした話者1間における対話進行を分析でき、さらに、この分析した二者毎の話者1間における対話進行に基づき、三者間における対話進行の過程の分析を好適に行うことができるといった、対話コーパスの作成に有用な三者会話データ収録装置Aを提供することができる。

10

【0038】

また、2チャンネル録音機3 a、3 b、3 cで異なる二者の音声を各チャンネルに分離して収録しているため、複数の話者が重複して発話した場合でも、収録した各話者の音声を聞き分けを確実にすることもできる。また、2チャンネル録音機を用いているため、簡単に本システムを構成することもできる。さらに、この録音部3で録音する音声を所定のサンプリング周波数でサンプリングし所定の値で量子化したデジタルデータ化とし、そして、タイムコードによって各音声の録音位置を好適に知ることができるため、例えば、会話コーパス作成のための転記作業を行う際において重複発話箇所の特定を容易に行うことができ、収録した音声データを好適に整理・分析することができる。

20

【0039】

そして、音響調整部5を、五台の音響ミキサを適宜接続して構成しているため、本システムを好適に構築することもできる。

【0040】

また、防音ボックスS内に話者1が入ることにより、他の話者の音声が混同しないように構成しているが、その防音ボックスS内の話者をビデオカメラVで撮影し、これを画像出力部8に表示させるとともに、他の防音ボックスS内の話者の音声を、この画像出力部8に映った方向と一致するようにして対をなす二つの音声出力ユニットから出力するように構成しているため、各話者に不自然さ等を感じさせず自然な対話コーパスの基となる会話を促すことができる。

30

【0041】

さらに、基準信号出力部から正弦波等の基準信号を出力することにより、前記録音部3の録音レベルや前記音声出力部の音声出力レベルを容易に調整できる。

【0042】

また、会話時の前記各話者の音声と各話者が映った映像とを収録する総合収録部を設けているため、対話コーパス作成時にこれを参照し、会話の全体を好適に把握することができる。

【0043】

すなわち、このような本発明の三者会話データ収録装置Aを用いれば、対話コーパスの作成に有用とすることができる。

40

【0044】

なお、本実施形態では、各話者1の音声A、B、Cを出力するためにヘッドホンを用いたが、インナホンやニアホンを用いてもよい。

【0045】

また、防音ボックスS内に話者1が入ることにより、他の話者の音声が混同しないように構成していたが、一の話者の音声が他の話者のマイクロホンに収録されなければ、防音ボックスを用いず通常の一部屋等で収録するようにしても構わない。

【0046】

さらに、防音ボックスS内の話者をビデオカメラVで撮影し、これを画像出力部8に表示

50

させるように構成していたが、例えば、防音ボックスSの一部が遮音ガラスで構成され各話者1同士がこの遮音ガラスを介して視認できる場合等、画像出力部8を使用しない実施態様も考えられる。

【0047】

また、各防音ボックスS内に2つの画像出力部8を配置したが、1つの画像出力部8に他の二者を映した映像を分割表示させる実施態様も考えられる。

【0048】

なお、本実施形態では、前記2チャンネル録音機3a、3b、3cの各チャンネルにおいて、各話者1の音声をサンプリング周波数が48kHz、16bitで量子化したデジタルデータで録音するように設定しているが、サンプリング周波数及び量子化レートはこれに限られるものではない。また、2チャンネル録音機をDATレコーダで構成したが、これに限らず、例えば、MDレコーダやDVDレコーダ等、デジタル録音可能なものであればよい。さらに、コンピュータのハードデスク等に音声を記録するといった実施態様も考えられる。

【0049】

また、本実施形態では、基準信号出力部6から所定の周波数の正弦波を出力することにより、前記録音部3の録音レベルや音声出力ユニットの音声出力レベルを調整するように設定していたが、基準信号出力部6から発生させる信号は、正弦波に限らず、ホワイトノイズやピンクノイズ等のランダムノイズやその他の信号であっても構わない。

【0050】

なお、音響調整部5を五台の音響ミキサで構成したが、音響調整部5を構成する台数はこれに限られない。また、音響調整部5を、所定のプログラムによってコンピュータを動作させて音響ミキサとしての機能を発揮するようにして構成するといった実施態様も考えられる。

【0051】

その他、各部の具体的構成についても上記実施形態に限られるものではなく、本発明の趣旨を逸脱しない範囲で種々変形が可能である。

【0052】

【発明の効果】

以上に詳述した本発明によれば、録音部に組みにして録音した二者毎の組にした話者間における対話進行を分析でき、さらに、この分析した二者毎の話者間における対話進行に基づき、三者間における対話進行の過程の分析を好適に行うことができるといった、対話コーパスの作成に有用な三者会話データ収録装置を提供することができる。

【図面の簡単な説明】

【図1】本発明の実施形態における三者会話データ収録装置を構成する音声系統図を示す図。

【図2】本発明の実施形態における三者会話データ収録装置を構成する映像系統図を示す図

【図3】同実施形態における三者会話データ収録装置における収録の実施態様を示す図。

【符号の説明】

A・・・三者会話データ収録装置

1(1a、1b、1c)・・・話者

2(2a、2b、2c)・・・收音部

3・・・録音部

3a、3b、3c・・・2チャンネル録音機

4aL、4aR、4bL、4bR、4cL、4cR・・・音声出力部(音声出力ユニット)

1d・・・第四の話者

6・・・基準信号出力部

7・・・総合収録部

10

20

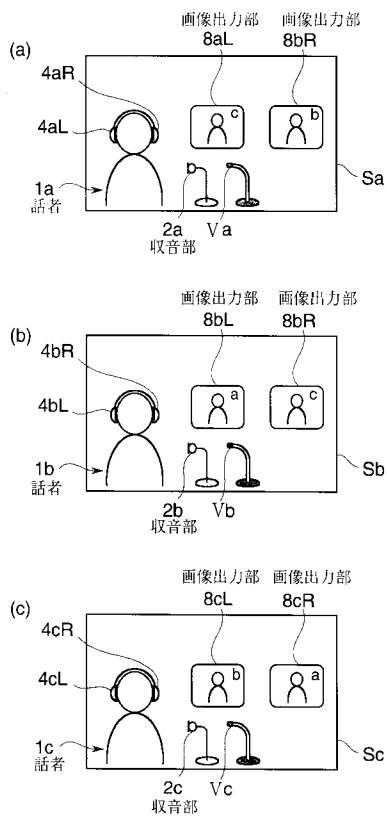
30

40

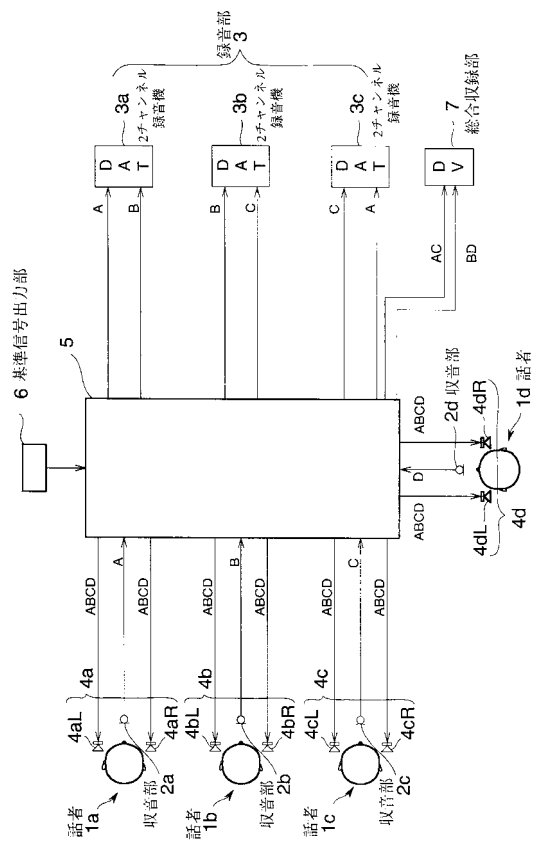
50

8 . . . 画像出力部

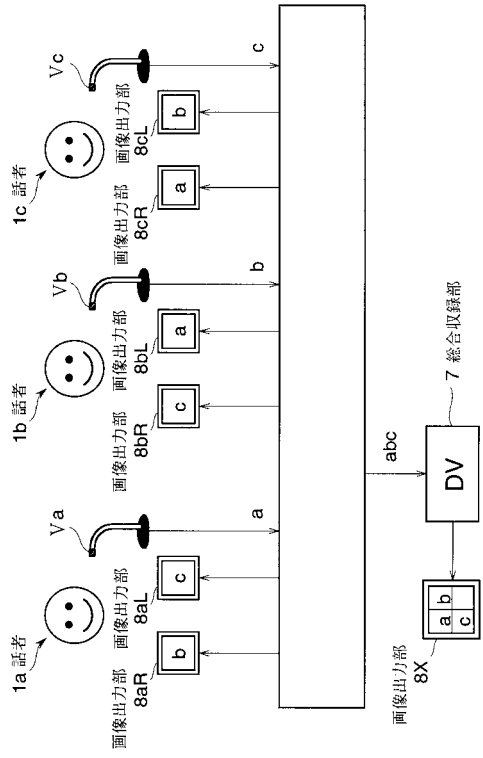
【 図 1 】



【 図 2 】



【 図 3 】



フロントページの続き

(51)Int.Cl.⁷

F I

H 0 4 R 3/00 3 2 0

(72)発明者 柴田 雅光

大阪府大阪市北区鶴野町3 - 10 株式会社インターグループ内

審査官 山崎 達也

(56)参考文献 特開平10 - 294798 (JP, A)

特開2003 - 060792 (JP, A)

特開2004 - 064236 (JP, A)

(58)調査した分野(Int.Cl.⁷, DB名)

G11B 20/00

G11B 31/00

H04N 7/15

H04R 3/00