

仮想の“音の部屋”によるコミュニケーション・メディア *Voiscape*

金田 泰

日立製作所 システム開発研究所

〒215-0013 神奈川県川崎市麻生区王禅寺 1099 番地

E-mail: kanada@sdl.hitachi.co.jp

あらまし 電話にかわるべきコミュニケーション・メディア *voiscape* の概念を提案する。Voiscape は、3次元オーディオ技術による仮想的な“音の部屋”をユーザ間で共有し、そのなかを自由に移動してさまざまなひとと会ったりわかれたりしながら多者間のコミュニケーションがおこなえるウェアラブルなメディアである。プレゼンスや周縁的情報の伝達を可能にし、電話におけるような1対1の会話から従来のメディアではできなかったさまざまなかたちのコミュニケーションまでをカバーすることによって、つながり感・安心感の共有や暗黙知の共有も実現されるだろう。この論文では *voiscape* の使用場面や手順についてのべ、PC 上に開発した *voiscape* のプロトタイプについてものべる。プロトタイプ上ではユーザは前方の様子を3次元グラフィクスによって確認しながらマウスをつかって部屋内を移動することができる。

キーワード 電話, IP 電話, 音声通信, リアルタイム通信, 3次元オーディオ, 3Dオーディオ, 多者間通話, Voiscape.

A Virtual “Sound Room” Based Communication-Medium Called *Voiscape*

Yasusi Kanada

Hitachi, Ltd., Systems Development Laboratory

Aso-ku Ozenji 1099, Kawasaki, 215-0013, Japan

E-mail: kanada@sdl.hitachi.co.jp

Abstract The concept of a new communication medium called *voiscape* is proposed. A virtual “sound room” that is based on the spatial audio technology is shared among the users in *voiscape*, and a person can move freely in the room, can meet and depart from other people, and can talk with two or more persons by using *voiscape*. By enabling transmission of presence and peripheral information, *voiscape* will cover from a telephone-style one-to-one conversation to a variety of communication types that are impossible in conventional media, and will enable sharing the feeling of connection and relief and sharing tacit knowledge. This paper describes usage scenes and a procedure of *voiscape* and also describes a PC-based prototype. In this prototype, the user can confirm the situation in front by 3D graphics, and can move around the room by using a mouse.

key words Telephone, IP telephone, Voice communication, Real-time communication, spatial audio, Multi-user communication, Voiscape.

1. 電話から仮想空間メディアへ

20世紀における科学技術の急速な発展にもかかわらず、電話の基本的なインタフェースは A. G. Bell が 1876 年に発明して以来、約 130 年間にわたってまったく変化していない。すなわち、電話によって会話するためには、まず相手に回線を接続し、1 個のマイクと 1 個のスピーカをつかって 1 対 1 で会話し、会話がおわったら回線を切断する。図 1 は 1878 年の電話器の例である (<http://www.atcaonline.com/phone/coffin.html>) が、これをみてもそれが変化していないことがわかる。携帯電話の普及は最近の 10 数年に電話システムをおおきくかえたが、この基本インタフェースはやはりなにも変化させなかった。

変化していないのは、けっしてこのインタフェースが理想的だからではない。たとえば、電話には接続するまで相手の様子がまったくわからないという問題点がある。そのため、相手が電話がくると非常にこまる時にも容赦なくかけてしまう。インタフェースが変化していない理由としてこのインタフェースがひとびとに支持されていることがあるのはまちがいないが、支持されているのはむしろ、このインタフェースにかわる有力なインタフェースが提案されていないからだとかんがえられる。インタフェースが変化していないおもな理由は電話のネットワークが回線交換網だからである。すなわち、

回線交換網を使用するかぎりには接続・切断は必須の操作であり、1 対 1 でしか会話できないという制約も容易になくせないからである。

電話網にはすでに、パケット交換の利点をとり入れた ATM などのネットワークが部分的に使用されてきているが、統一されたインタフェースによってどこでもだれでもアクセスできるという電話の利点をいかすためには、上記のような回線交換にもとづくインタフェースを変更するわけにはいかなかった。しかし、いまや、IP ネットワークによって電話網は急速に置換されようとしている。固定電話はすでに IP 電話によって侵蝕されているが、携帯電話もいずれ IP 化することはまちがいないであろう。IP ネットワークにおいてはパケット交換を使用しているため常時接続が基本であり、会話しないときにもすこしだけパケットをおくりつづけることができる。電話網がやわらかい IP ネットワークによって置換されれば電話のインタフェースを制約していた原因も消滅し、直接のコミュニケーションに匹敵する、人間の聴覚能力を

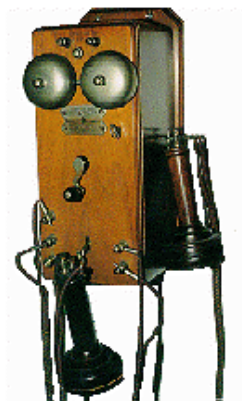


図 1 1878 年の電話器

いかした、3人以上での自由な会話ができる、常時接続を前提としたあたらしいメディアによって電話はとってかわられるであろう。

それでは、従来の電話にかわるインタフェースはどのようなものだろうか。そのこたえをみいだすために、もう一度電話についてかんがえる。電話は遠隔地にいるひとと、あたかもすぐそばにいるかのように話ができるようにするメディアである。これは、電話によって仮想空間がつくられていることを意味する。つまり、電話の用途が基本的に1対1の会話に限定されているためにふだんは気がつかないものの、本質的に仮想環境をつくるためのメディアだということの意味している。すなわち、仮想空間はコンピュータ・ゲームなどにおける仮想環境に固有のものではなく、普遍的なものである。この仮想空間の存在は、たとえばつぎのような大学生の話によって例証されているとかがえられる。

長電話をして2時間、3時間喋っていると、自分の所在がわからなくなる様な不思議な錯覚に陥ることがある。電話での話題は、実際に自分がいる場所や時間とは異なる時空を持っている。……そこで展開する話に没頭しているうちに、いわばその世界にトリップしてしまい、ふと我に帰ってみて現実の自分の所に違和感を感じる。(大学3年の女性の話)[Yos 91]

ここで“実際に自分がいる場所や時間とは異なる時空”と表現されているものは、上記の仮想空間とおなじものだとかがえられる。この話におけるような“錯覚”がおこることはまれだが、“錯覚”がおこるかどうかは、仮想空間がどれだけの現実性をもって感じられるかのちがいだとかがえられる。

2. 電話にかわるべきメディア voiscap

この章では電話にかわるべき音声コミュニケーション・メディアのすがたをうきぼりにしたい。作曲家 Murray Schafer が音によってつくられる風景を soundscape とよんだ [Mur 77] のにならって、このメディアを voiscap (声の風景、声景) とよぶことにする。

2.1 常時接続型環境の実現

メディアをとおさない直接の会話においてはプレゼンスや周縁的(peripheral)情報がつたえられ、それらがコミュニケーションにおいて重要なやくわりをはたしているとかんがえられる。プレゼンスとは、会話の相手となるべきひとや操作対象のものが現前していることをメディア使用者が感じる感覚のことである。また周縁的情報とは、会話そのものではなく会話の相手やその周辺の事物に関する会話の周辺の情報のことである。ここではそれらのやくわりと、メディアにとりいれる可能性について検討する。

第1に、直接の会話においては相手がちかくにいればプレゼンスを把握することができる。つまり、会話をしていないときでも相手のようすがわかり、会話できる状態かどうか把握できる。また、すでに他者と会話していてもそれをいちじるしくじゃますることなく会話を始められることがおおい。また、家族や仲間のプレゼンスがつたわることで、つながり感や安心感がえられるであろう。電話をはじめとする旧来のメディアをとおした会話においては、回線を切断しているあいだはプレゼンスを把握することができない。つまり、会話していないときは相手の状態がまったくわからず、相手の都合がわるいときに会話を強制するばあいがある。

この点に関して、インスタント・メッセージング(IM)などにおいては接続・切断の機能は従来の電話と同様にして、そのかわりにこれらのメディアを補足するものとしての狭義の“プレゼンス” [Day 00] を導入されている。ここで“プレゼンス”とは、ひとことでいえばメディア通信していない状態でもつたえられる相手に関する情報のことである。たとえば、いま電話のちかくにいるかどうか、電話にでられる状態かどうか、などといった情報のことである。

しかし、“プレゼンス”をつたえるために記号的なプロトコルがつかわれているので、現在“プレゼンス”とよばれている情報はほとんど機械的に解釈され記号化された情報にかざられている点に問題点がある。すなわち、記号化する際に重要な情報がうしなわれたり、まちがった解釈がなされる可能性がある。

また、IMなどにおいては“プレゼンス”と会話とが分断され、ことなるしかけで伝達される。しかし、常時接続が基本のインターネットのもとではそもそも意識的な接続・切断じたいが不要かつ有害であり、会話とプレゼンスを分断することによってあらたな複雑さ・不自由さもちこむ必要はないであろう。たとえばインターネット上のビデオ電話においては相手の映像は会話の一部であり、相手が電話にでられるかどうかをしめす情報はプレゼンスだとかがえられるが、これらはいずれも相手の様子に関する情報であり、それらのあいだに本来は境界は存在しない。映像や音声は記号化するのが困難なプレゼンスを、部分的ではあるが、つたえられる。

第2に、直接の会話においては音響、臭気をはじめとする記号化されていない情報もふくめて、さまざまな周縁的情報がつたわる。これらのなかにはコミュニケーションと無関係なものもおおいが、その文脈を形成するものもある。つまり、周縁的情報がないと会話の内容だけをきいても文脈が理解できないばあいがある。これに対して従来のメディアをとおした会話においては周縁的情報はつたわりにくい。電話においては回線が切断されていればいっさいの情報がつたわらないから、切断時にはもちろん周縁的情報がつたわらない。接続時には話者の周辺で発生する音声や周縁的情報としてつたわるが、非常に限定的である。狭義の“プレゼンス”が存在するときは切断時にもある程度の周縁的情報がつたわるばあいがあるが、つたわるのはかざられた情報だけである。

Voiscapにおいてはこれらの問題を解決するために、IPネットワークの利点をいかして常時接続型の環境を実現するべきであろう。ただし、通信のために常時広帯域を使用すると光ネットワークにおいても容量が不足するとかがえられるので、会話していないときには帯域をおさえるくふうが必要であり、そのために回線の自動接続・切断の機能も必要になるであろう。

2.2 仮想の会話空間の共有

この報告の冒頭でのべたように遠隔通信を可能にするメディアは本質的に仮想空間を形成するが、この空間は会話者間で共有する必要があるとかんがえられる。たとえば、AとBとの会話においてAがBの近傍にいると感じているのにBはAが遠方にいると感じていると、会話はうまくいかないであろう。また、AはBの話をきくことができるのにBはAの話をきくことができないというような非対称な状況をなくさなければ、このメディアを安心してつかうことはできないだろう。会話空間の共有はとくに3人以上での会話において重要だとかがえられる。したがって、voiscapにおいてはこの仮想空間を演出し、ユーザが仮想空間の存在を意識する必要があるだろう。この空間はかならずしもユークリッド空間である必要はないが、人間の直観や感覚がよくはたらくようにするためには、3次元または2次元のユークリッド空間を使用し、その音響特性も現実の空間にちかいものにするのがよいとかんがえられる。すなわち、ここに3Dオーディオ技術を導入するべきであろう。

Voiscapにおける仮想空間のイメージを図2に示す。ここには部屋とよばれる複数の仮想空間があり、そのなかから1個を選択することができる。ただし、部屋はかならずしもグラフィクスなどを使用してユーザに視覚的にみせるのではなく、聴覚的にのみ存在させることも可能である。部屋が聴覚的に存在するということは、残響はもちろん、ばあいによっては壁による反射も計算するべきだということの意味している。IMなどでも多者間通信のために部屋のようなも

のが使用されることがあるが、voiscape における部屋はこれを拡張してさまざまな属性を追加したものだということができる。ユーザは部屋内を自由に移動できる。声のおおきさ、明瞭度などの“通信情報量”は部屋内での距離により単調減少するのが自然である。

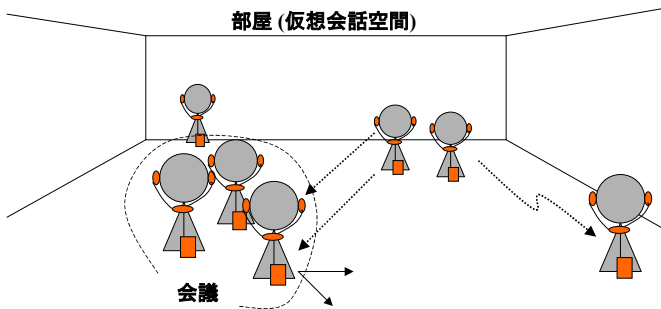


図2 仮想空間のイメージ

Voiscape においては、この空間のなかで各参加者は自由に自立的に移動できる必要がある。この空間およびそのなかでの参加者の位置や距離が、すべての参加者について総合的であることが重要だとかんがえられる。そのためには、各参加者が仮想空間の属性(たとえば部屋のおおきさ、距離による音声の減衰特性、残響特性)などを共有する必要がある。また、部屋のなかでの参加者の位置をはじめとする各種の属性も共有する必要がある。

2.3 多者間の複数の会話の支援

電話においては基本的に会話が1対1でおこなわれる。パーティラインを使用すれば3人以上での会話も可能だが、そのばあも基本的に会話の開始から終了まで参加者が固定され、途中から参加するのは困難である。そのため、電話においては会話中は相手を独占し、相手が他の会話に参加するのを困難にしている。会議システムにおいても、そのおおくは会議参加者を集中的に管理するため、自由な参加・退出が困難なもののおおい。また、おおくは会議室における会議を代用するものであって、特定の場所を特定の時間に予約して使用しなければならない。

これに対して直接のコミュニケーションは、電話のように1対1であることは強制されないので、多者間会話が基本だとかんがえることができる。会話の途中で自由に会話に参加したり退出したりすることができる[She 95]。Voiscape においても、自由な会話のために1対1という制約はなくすべきである。多対多を基本とするところから、voiscape は基本的に会議システム(conference system)である。会議参加者の管理法をくふうして、すでにおこなわれている会話にあとから参加したり、会話から退出したり、会話グループが自立的に分割・融合したりすることが自由にできるようにするべきである。しかも、このような会話の再編成のためにできるだけ意識的な操作をせずにすむようにするのがよい。

また、電話や従来の会議システムにおいては音声がモノラルでつたえられるため、距離も方向もわからない。そのため、1対1のときはよいが、3人以上での会話においては話者を特定するのが困難であり、また複数の話者が一度に会話すると、それらを分離するのが困難である。電話における音声の明瞭性が周囲の雑音に影響されやすいひとつの原因も距離感・方向感の欠如にあるとかんがえられる。携帯電話はしばしば騒々しい環境のなかで使用されるが、このような騒音がおおきい環境においてはバイノーラル音声(両耳聴音声)の使用が有効だと報告されている[Abo 01]。

これに対して直接のコミュニケーションにおいては音声に距離感・方向感がある。これらは音声そのものからもある程度わかるが、さらに視覚的に補強される。そのため、話者の特定は容易であり、また複数の会話の分離が容易になる、すなわちカクテル・パーティ

効果 [Che 53] がえられる。このような人間の聴覚能力をいかすためにも voiscape には3Dオーディオ技術を導入するべきであろう。

2.4 一方的な“会話”の実現

会話は本来、双方向のものであり、その本性をメディア上で実現したのが電話であることはまちがいない。しかし、電話における会話がこの形式にしばられているのに対して、直接のコミュニケーションにおける会話はかならずしもこの形式にとらわれない。すなわち、会話に参加していない第3者が会話を一方的に聴講できることがしばしばある。また、会話に参加していないひとに、会話への参加をうながさないうまま、きこえるように話をすることもできる。前者は一方的な聴講であり、後者は一方的な発言とよぶことができるだろう。

このような一方のコミュニケーションは部分的には従来から一方方向性のメディアにおいて実現されている。とくに、テレビやラジオは一方的な視聴または聴講のためのメディアとして普及している。一方的な視聴はメディアへの人間の受動的な関与を可能にする。電話はその双方向性のために、また音声以外の情報がつたわらないために、会話中は非常に能動的であることを話者に要求する。能動的な状態を維持するためかなりの“エネルギー”が必要とされ、会話に専念することがもとめられる。これに対してテレビやラジオは受動的に視聴できるのでくつろぐことができ、他の作業をしながら視聴することができる。

Voiscape においては上記のような一方的な“会話”を可能にするべきであろう。たとえば、この空間内に身をおきながら、空間内の会話に積極的に関与はせず、それにじゃまされずに、しかも“つながったまま”単独で作業することもできるであろう。“つながったまま”とは、作業中も部屋内の重要な情報がつたわり、なにげなく情報はいってくるということの意味している。テレビのようなマスメディアとはちがって、voiscape はこれまでのメディアになかった小規模な一方的な会話を実現するあたらしいタイプのメディアである。

上記のような一方的な“会話”は知識マネジメント(knowledge management) [Non 95] において強調されるコミュニティにおける暗黙の知識の流通におおきなやくわりをはたせるであろう。J. S. Brown はある企業のサービスセンターにおけるベテランと新人という2人のコールセンターのオペレータのあいだの一方的な“会話”が新人をベテランに匹敵するまでにそだてあげた例をあげている[Bro 02]。このような暗黙知のはたすやくわりは、けっして企業活動に特有のものではなく、創造的な活動すべてに共通のものだとかんがえられる。

2.5 セキュリティとプライバシー保護

Voiscape においては多者間会話や一方的な会話の導入などによって電話にくらべてプライバシー侵害の危険がふえているので、その保護に配慮するべきである。また、セキュリティに関しても配慮しなければならないことはいうまでもない。

セキュリティに関しては、voiscape の外部からの保護と、複数の部屋を使用するときの部屋間の保護に関しては暗号化の技術を使用すればよい。部屋内は基本的にセキュリティ上の問題があるユーザは排除されているとかんがえられるので、ここでは部屋内のセキュリティについてはかんがえない。

以下、プライバシー保護についてかんがえる。直接のコミュニケーションにおいては音声が周囲にもれないように会話したり、会話参加者のなかの一部のひととだけと会話したりすることができる。すなわち、さまざまなレベルのプライバシーを選択することができる。これに対して電話においては、多者間の会話においてもプライバシーのレベルを制御するのは困難である。相手と接続するまえにシグナリングによってある程度の設定ができる可能性はあるが、接続したあとは非常にかぎられた操作しかできない。

Voiscapе においてはプライバシーのレベルを自然な方法で選択できる必要があるだろう。そして、そのために仮想空間内の距離などによって通信や表示を制御するべきであろう。たとえば部屋内で近接しているひとの声ははっきりきこえるが、遠隔のひとの声はたとえ増幅してもききとれないようにするのがよいだろう。そのためには、遠距離の相手とは通信しないか、または通信するとしてもなんらかのひずみをいれる必要がある。たとえば、音量がおおきききこえただけ送信する方法や、声を他の音声で置換して送信する方法がかんがえられる。このような制御によってユーザのプライバシーが保護されると同時に、通信量が削減できる。このような制御をポリシーを使用して実現する方法について金田 [Kan 93] が記述している。

2.6 携帯性

Voiscapе においては携帯性が重要だとかんがえられる。それは携帯電話のかわりに使用するためにも必要だが、それだけでなく、常時使用することができるためにはウェアラブルである必要があり、そのためには携帯可能でなければならないという理由からである。デスクトップの端末を使用すると、つくえをはなれるたびにヘッドセットをはずさなければならないのでわずらわしいうえに、常時接続性がうしなわれる。したがって、voiscapе の端末としては、たとえば図 3 のようなものが想定される。ただし、現在のマイクつきイヤフォンやヘッドセットにはワンタッチで装着できないなどさまざまな問題があるので、それを解決する必要がある。

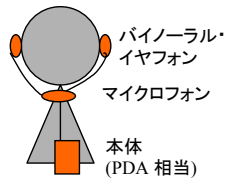


図 3 端末のイメージ

携帯電話のように実空間を移動しながらでも使用できるようにするには、常時ディスプレイをみなくても会話できる必要がある。ヘッドマウント・ディスプレイを使用することもかんがえられるが、移動時にもそれに依存するのは安全上問題がある。

3. Voiscapе における会話の手順

この章では voiscapе における会話のための手順についてのべる。Voiscapе においては複数のサーバと複数のプロトコルを使用するが、ここではそれらをくべつせず、一括して“サーバ”とよぶ。

まず、詳細な手順をしめすまえに概要をしめす。クライアントを起動すると自動的にサーバにログインし、サーバからおくれた入室可能な部屋のリストが表示される。ユーザがそのうちのひとつを選択すると、その部屋に入室している他のユーザとのあいだで通信が開始され、部屋の様子が 3D オーディオによる声景 (voiscapе) とグラフィクスによって表現される。ユーザは部屋のなかを自由に移動し回転することができる。3D オーディオを使用しているため、移動して話者にちかづけば声はおおききこえるし、回転すれば声の方向が変わる。

3.1 ログインと部屋リストの表示・選択

クライアントにはあらかじめユーザ名が入力されている。クライアント起動時にこのユーザ名をつかってサーバに対して自動的にログイン操作がおこなわれる。4 章で説明するプロトタイプクライアントは Windows PC 上で動作する (したがって、現在、携帯性は実現されていない) が、そのクライアントのウィンドウを例として図 4 にしめす。

ログインするとサーバから部屋リストが送付される。図 4 においてはウィンドウの左側にそのリストを表示している。Office、Project-X、

Meeting room、Home という 4 つの部屋名が表示されている。このユーザは通常の仕事を (このメディアを使用しながら) する際には同僚などが登録メンバーとなっている Office という部屋を使用し、その仕事とはべつプロジェクト X の仕事をする際にはその仕事におけるなかまが登録メンバーとなっている Project-X という部屋を使用し、部外者などとの会議をおこなうときには MeetingRoom という部屋を使用し、家族との連絡をとるときには MyHome という部屋を使用することを想定している。

ユーザは部屋リストのなかから部屋を選択して入室する。現在は一度に入室することができる部屋はひとつだけである。したがって、他の部屋を選択して入室するとそれまで入室していた部屋からは自動的に退室する。また、すべての部屋から退室して、全通信を終了させることもできる。

3.2 入室時の通信と部屋の表現

ユーザが部屋を選択して入室すると、部屋内にだれがいるかがわかるようになる。部屋内の他のユーザとのあいだで自動的に音声通信が開始され、仮想空間に位置づけられた他のユーザが 3D オーディオ / グラフィクスによって表示される。ヘッドセットを通じて 3D 音声ユーザの両耳につたえられる。

人間は音声だけで話者がいる方向を的確に把握することができない。とくに、聴覚による前後や上下のくべつは 4 章で説明する HRTF を使用してもなおあいまいである。そこで、オーディオ表示とあわせて、ユーザの周囲の様子をグラフィクスによって画面上に表示する。図 4 においてはユーザの前方が表示されている。ユーザ自身は表示されず、部屋は床と壁によって表現されている。ユーザ間の通信は音声だけにかぎられるので、ここでは他のユーザの画像は表示せず、直方体と円錐とをくみあわせて他のユーザを表現している。これだけでは他のユーザのむきがわからないが、直方体の上部にユーザ名を表示して、それがだれであるかがわかると同時に他のユーザの方向がわかるようにしている。

3.3 部屋内での移動と回転

ポインティング・デバイスを使用することによって、部屋のなかで自由に移動したり、方向をかえたりすることができる。この移動や回転は仮想空間内のものであるから、基本的に実世界における移動や回転とは無関係である。携帯性を実現するには他の適切なデバイスを選択する必要があるが、プロトタイプにおいてはポインティング・デバイスとしてマウスを使用している。マウスの左ボタンによって前後にドラッグすれば前後に移動できる。また左右にドラッグすれば左右にむきをかえることができる。ひとが実空間内を移動するときは、通常、前方への歩行とむきの変更とによって移動するので、それに対応する操作ができるようにしている。移動にもなって自動的に他のユーザとの通信を開始したり終了したりし、声がかきとれないようにしたりすることは 2.5 節においてのべたとおりである。

ポインティング・デバイスを使用して部屋内を移動するとき、移動を検出するごとにそのクライアントのグラフィクス表示を更新す



図 4 クライアント・ウィンドウ

る。それは、移動にともなう位置や方向の変化はすみやかにユーザにフィードバックするのがよいからである。しかし、サーバや通信相手にはこの変化をつたえる頻度をおとすくふうをしている。

4. プロトタイプ

この章においては、試作した voiscaple プロトタイプの概要とクライアント(端末)の実装についてかんたんに説明する。図5がプロトタイプの全体構成である。プロトタイプはおおきくわけるとサーバ群と複数のクライアントとで構成されている。現在、クライアントは Microsoft Windows XP または Windows 98 を搭載した PC 上で動作させている。

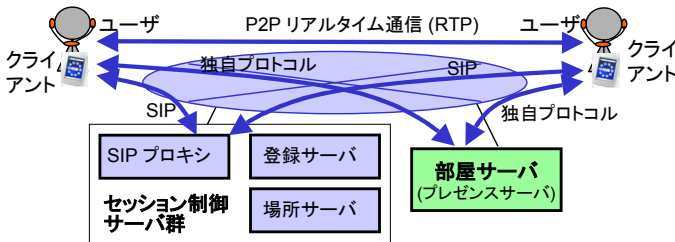


図5 試作したプロトタイプの全体構成

サーバ群は、セッション制御のためのサーバ群と voiscaple に特徴的なプレゼンスサーバ(部屋サーバ, Presence Server / Room Server)とで構成されている。セッション制御には SIP (Session Initiation Protocol) [Ros 02] の旧版 (RFC 2543) を使用している。プレゼンスサーバはクライアントがネットワークに接続されているかどうかなどのプレゼンス情報を管理し、voiscaple における仮想会話空間すなわち部屋とその利用者を管理する。SIPサーバに関してはとくに新規な点はないが、プレゼンスサーバの実装については金田 [Kan 93] がのべている。

クライアントの構造を図6に示す。端末は入力デバイスとしてマイクと位置指定デバイス(現在はマウス)をもち、出力デバイスとしてイヤホンまたはヘッドフォンとディスプレイをもち、もちろんマイクとヘッドフォンをあわせたヘッドセットを使用することもできる。マイクからの出力は音声入力部によってデジタル信号にエンコードされる。その信号は通信相手が存在するときには音声通信部におく。現在は 8000 Hz でサンプリングし、ITU-T G.711 u-law 64 kbps の信号として RTP (Real-time Transport Protocol) [Sch 96] によって他のクライアントに P2P で送信する。音声のキャプチャと再生、RTP による送受信には Java の拡張 API である JMF (Java Media Framework) [Gor 98][Fai 00] を使用している。

部屋モデラにおいては、位置指定デバイスからの出力にもとづいてユーザの部屋内における位置をもとめて、独自プロトコルによってプレゼンスサーバに送信する。また、プレゼンスサーバから他のユーザの部屋内における位置を受信する [Kan 93]。これによって 2.2 節でのべた仮想空間の共有を実現する。

3次元オーディオ/グラフィクス表示部においては、受信した RTP 信号を部屋モデラからえられる位置情報にしたがって 3次元音場に

位置づける。左右だけでなく前後や上下方向もあわせた方向感を表現するため、HRTF (Head Related Transfer Function) [Beg 00] 機能を取り入れた。HRTF とは人間の頭部周辺の音響特性のことである。この技術が 3D オーディオ技術の核だとかんがえられる。また、受信した信号に対応するユーザを 3D グラフィクスによって表示する。この表示には Java の拡張 API である Java 3D [Sow 00] を使用している。Java 3D においては 3D オーディオ表示機能も提供されているので当初は JMF とくみあわせればほぼ目的のプログラムが実現できると予想していたが、Java 3D はそのままではリアルタイム通信とくみあわせられないことがわかり、独自にオーディオ API (図5の JA3D) を開発して Java 3D と併用した。

セッション制御部においては SIP を使用し、SIP プロキシを経由して他のクライアントとのあいだの RTP 通信の開始・終了等を制御する。ポリシー制御部はおもにセッション制御部の機能を制御するためのポリシーを保持し、それにもとづいて 2.1 節や 2.5 節でのべた回線の自動接続・切断を実現している [Kan 93]。

5. 試作の結果と検討

プロトタイプにおいてリアルタイム通信と 3D オーディオ/グラフィクスをくみあわせて動作させることには想像以上に困難があった。上記の Java 3D と JMF とのくみあわせの問題のほか、音質劣化・遅延 [Kan 93] などの問題が発生した。未解決の問題もあるため、まだユーザによる試用が可能な状態ではない。したがって、ここではプログラム開発の過程やシステム・テストにおいてわかったユーザ・インタフェースに関する課題についてだけのべる。

- 部屋内の移動・回転に関する課題: 実空間においてはほとんど努力なしに無意識的に空間内を移動できるのに対して、プロトタイプにおいては部屋内の移動や回転のためにユーザはある程度の努力をもとめられる。Voiscaple の目的はコミュニケーションであり、部屋内の移動はそれを補助するためのものなので、この努力はへらす必要がある。努力が必要な一因はマウスの自由度がおおきすぎることである。プロトタイプにおいては移動を平面内にかぎって自由度をへらしているが、それでも操作はそれほど容易ではない。この問題を解決する一方法は、原始的だが、マウスやトラックボールのような連続的な入力デバイスをつかうかわりにカーソルキーのような離散的なデバイスを使用することである。とくに、正確に連続的な入力をするのが困難な老人や身体障害者などにおいては、離散的な入力のほうが適切であろう。カーソルキーならば低コストであり、PDA や携帯電話にすぐ実装できる。

- ポインティング・デバイスに関する課題: 現在のプロトタイプにおいてはマウスを使用しているが、自由度がおおきすぎるという前記の問題以外にもマウスには 2 つの問題点がある。第 1 に、マウスをつかうと実空間の移動時に携帯電話と同様に使用できるようにするのが困難になる。この問題はマウス以外の適切なポインティング・デバイスを使用すれば解決できるだろう。第 2 に、マウスでは一定方向には一定量しか移動できない。たとえば、ポインタが画面の最上部に達するか、またはマウスがマウスパッドの最前部に達すると前方に移動できなくなる。さらに移動するには、いったんボタンをはなすか、マウスをもちあげて移動させる必要がある。

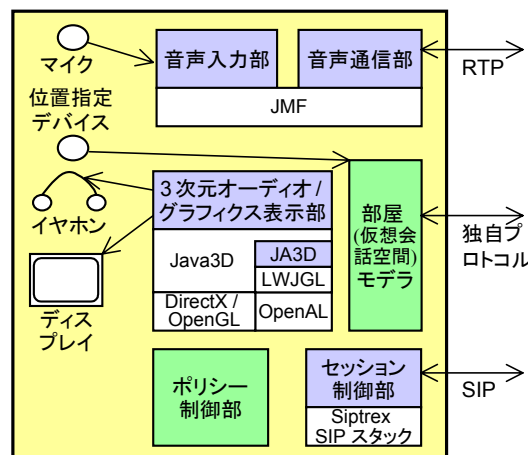


図6 クライアントの構造¹

¹ DirectX (Microsoft の登録商標) と OpenGL (Silicon Graphics 社の登録商標) はグラフィクス API, OpenAL はオーディオ API, LWJGL (Light-Weight Java Game Library) は SourceForge.net において開発された、Java から OpenGL, OpenAL を使用するための API である。Siptrex は UCL (University College London) において開発された SIP スタックとアプリケーションの名称である。

この問題はマウスホイールやジョグダイヤルなどのデバイスを使用すれば解決できる。

- 音像定位に関する課題: プロトタイプに使用した 3D オーディオ技術は低コストのものであり、最高の技術ではない。また、音像をきちんと定位させるためには、本来は個人ごとに特性を調整する必要があるが、調整できない。そのため方向感はかなりあいまいであり、距離感も不十分である。したがって、今後、PC サウンド以外の方法も検討する必要がある。

6. 関連研究

DIVE (Distributed Interactive Virtual Environment) [Ben 93] は共同作業 (CSCW) などのための仮想環境である。仮想空間上のオブジェクト (人間もふくむ) はその周囲にオーラ (aura) とよばれる領域をもち、それが他者のオーラとかさなると会話などの相互作用がうまれるというモデルにもとづいている。相互作用はあるかないかの 2 とおりであり、そのつよさはモデル化されていない。

MASSIVE [Gre 95] は共同作業などのための会議システムである。DIVE と同様の空間的なモデルを採用し、オーディオ、グラフィックス、テキストなど、さまざまなメディアをつかった会議ができる。相手との距離がひろがると相互作用がよくなるという *voiscape* にちかいモデルを採用している。ひとつの空間内で複数の会議をひらくことができ、そのあいだで相互作用がありうることも指摘している。MASSIVE3 にいたる 3 世代のシステムが開発されている。

京大の中西ら [Nak 98] は多対多の非形式的なコミュニケーションをサポートするため、仮想空間内を自由に移動しながら会話することができるシステム *FreeWalk* を開発している。仮想空間をワークステーション上の 3D グラフィックスによって表示し、ビデオをテキストチャマップによってはりつけた四角錐によって話者を表現している。音声も距離にしたがって減衰するようにしている。

Interval Research 社の Singer らは *Somewire* [Sin 99] というシステムを開発している。長時間にわたる仕事上の会話を支援するには常時接続を基本とする会議システムが必要だと主張し、いくつかの軽量な会議システムを開発し使用実験をおこなっている。その結果、ビデオは不要であり、高品質のステレオ・オーディオ (3D オーディオではない) だけのシステムがよいと結論している。対話相手の位置が制御できる、GUI より直観的な物理的 (tangible) インタフェースをためしているが、相手の位置を自由に制御できるようにすることはユーザの混乱をまねくと結論している。2.2 節でのべた仮想空間の共有が重要であることをしめしているといえるだろう。

MIT の Rodenstein らは音声による会議を (仮想) 空間に位置づけて (grounded) おこなう環境を開発している [Rod 00]。音声は距離にしたがって減衰させている。空間内で接近して会話したり、会話のサブグループを形成したりするできるようにすることを意図している。共同作業を目的としたシステムではなく、社会的コミュニケーションを意図している。

7. 結論

電話にかわるべきコミュニケーション・メディア *voiscape* のあるべきすがたについて考察し、会話の手順をしめすとともに開発したプロトタイプについてのべた。考察によって、リアルタイム通信の技術と 3 次元オーディオ/グラフィックスの技術をくみあわせ、人間のコミュニケーション能力をいかに IP ネットワークの常時接続性をいかにしたメディアのすがたをえがくことができたとかんがえられる。また、プロトタイプ開発によってさまざまな課題が抽出された。今後、現在のプロトタイプにのこされた課題を解決したい。また、このプロトタイプはまだ 2 章でのべた要件の一部しかみたしていないので他の要件もとりいれ、そのうえで認知的な評価をおこないたい。

参考文献

- [Abo 01] Abouchacra, K. S., Breitenbach, J., Mermagen, T., and Letowski, T., “Binaural Helmet: Improving Speech Recognition in Noise with Spatialized Sound”, *Human Factors*, Vol. 43, No. 4, pp. 584–594, 2001.
- [Beg 00] Begault, D. R., “3-D Sound for Virtual Reality and Multimedia”, NASA/TM-2000-XXXX, NASA Ames Research Center, April 2000, http://human-factors.arc.nasa.gov/ihh/spatial-papers/pdfs_db/Begault_2000_3d_Sound_Multimedia.pdf
- [Ben 93] Benford, S. D., and Fahlén, L. E., “A Spatial Model of Interaction in Large Virtual Environments”, *3rd European Conference on CSCW (ECSCW'93)*, Milano, Italy, Kluwer, 1993.
- [Bro 02] Brown, J. S. (宮本 喜一訳), “なぜ IT は社会を変えないのか”, p. 164, 日本経済新聞社, 2002-3.
- [Che 53] Cherry, E. C., “Some Experiments on the Recognition of Speech, with One and with Two Ears”, *Journal of the Acoustical Society of America*, Vol. 25, pp. 975–979, 1953.
- [Day 00] Day, M., Rosenberg, J., and Sugano, H., “A Model for Presence and Instant Messaging”, RFC 2778, IETF, February 2000.
- [Fai 00] Faiman, N., Giese, D., Rokanuzzaman, A., and Schroeder, M., “A Survey of the Java Media Framework 2.0”, CSci 532 : Programming Languages and Paradigms, University of North Dakota, <http://www.cs.und.edu/~mschroed/cs532/survey.doc>.
- [Gor 98] Gordon, R. and Talley, S., “Essential JMF – Java Media Framework”, Prentice Hall PTR, November 1998.
- [Gre 95] Greenhalgh, C., and Benford, S., “MASSIVE: a collaborative virtual environment for teleconferencing”, *ACM Transactions on Computer-Human Interaction (TOCHI)*, Vol. 2, No. 3, pp. 239–261, September 1995.
- [Kan 93] 金田 泰, “仮想の‘音の部屋’によるコミュニケーション・メディア *Voiscape* におけるポリシーベース・セッション制御”, 電子情報通信学会技術研究報告 (IA/IRC/QAI 研究会), 2003-10-8.
- [Mur 77] Murray Schafer, R., “The Tuning of the World”, 訳書: 鳥越 けい子他訳, “世界の調律”, 平凡社, 1986.
- [Nak 98] 中西 英之, 吉田 力, 西村 俊和, 石田 亨, “FreeWalk: 3 次元仮想空間を用いた非形式的なコミュニケーションの支援”, 情報処理学会論文誌, Vol. 39, No. 5, pp. 1356–1364, 1998.
- [Non 95] 野中 郁次郎, 竹内 弘高, 知識創造企業, 東洋経済新報社, 1996.
- [Rod 00] Rodenstein, R., and Donath, J. S., “Talking in Circles: Designing A Spatially-Grounded Audio Conferencing Environment”, *ACM CHI 2000*, pp. 81–88, April 2000.
- [Ros 02] Rosenberg, J., Schulzrinne, H., Camarillo, G., Johnston, A., Peterson, J., Sparks, R., Handley, M., and Schooler, E., “SIP: Session Initiation Protocol”, RFC 3261, IETF, June 2002.
- [Sch 96] Schulzrinne, H., Casner, S., Frederick, R., and Jacobson, V., “RTP: A Transport Protocol for Real-Time Applications”, RFC 1889, IETF, January 1996.
- [She 95] Shenker, Weinrib, and Shooler, “Managing Shared Ephemeral Teleconferencing State: policy and Mechanism”, draft-ietf-mmusic-agree-00.txt, Internet Draft, IETF, July 1995.
- [Sin 99] Singer, A., Hindus, D., Stifelman, L., and White, S., “Tangible Progress: Less Is More In Somewire Audio Spaces”, *ACM CHI '99*, pp. 104–112, May 1999.
- [Shi 01] Shinn-Cunningham, B., “Creating Three Dimensions in Virtual Auditory Displays”, In *Usability Evaluation and Interface Design: Cognitive Engineering, Intelligent Agents and Virtual Reality*, M. J. Smith, G. Salvendy, D. Harris, and R. J. Koubek eds., Erlbaum, pp. 604–608, August 2001.
- [Sow 00] Sowizral, H., Rushforth, K., and Deering, M., “The Java 3D™ API Specification (2nd Edition)”, Addison-Wesley, May 2000.
- [Yos 91] 吉見 俊哉, “個室のネットワーク — 電話コミュニケーションと生活空間の変容”, 東京大学新聞研究所紀要 43 号「電話コミュニケーションの現在」, 1991.