社団法人 情報処理学会 研究報告
IPSJ SIG Technical Report

2006－DPS－126 (39)
2006－CSEC－ 32 (39)
2006／3／17

視覚情報により強化された3Dサウンド場による共有型多人数音声チャットシステムの設計と実装

大橋純 広渕崇宏 河合栄治 藤川和利 砂原秀樹

奈良先端科学技術大学院大学 情報科学研究科

概要

立食パーティのようにコミュニケーション空間に複数の話題が存在することが可能である「場」共有型のリアルタイムコミュニケーションシステムを提案し実装した。提案システムは、3Dサウンドを利用することでコミュニケーション空間に距離と方向の概念を導入し、現実世界と同じく複数の会話が並存できる環境を実現した。会話の音源を既存の環境で聞き取りやすい位置に移動し、視覚情報を利用して音源定位を補助するというアプローチで、複数の会話が並存する際に会話の識別が難しいという問題の解決を図った。評価により提案手法の有効性を示した。

Design and Implementation of Multiuser Voice Chat System Applying 3D Sound Space Emphasized by Visual Information

Jun Ohashi Takahiro Hirofuchi Eiji Kawai Kazutoshi Fujikawa Hideki Sunahara

Graduate School of Information Science, Nara Intsitute of Science and Technology

Abstract

We propose and implement "Space Sharing" multiuser voice chat system which can contain parallel conversations in the same communication space. To realize parallel conversatoins, the proposed system uses a concept of distance and direction at communication space using 3D sound. Limitation of Sound Devices and 3D sound libraries make it difficult to distinguish each conversation at same communication space. To resolve this problem, we propose the way to move sound sources to readily-observable position and emphasize source localization by visual information. We verify the effectivity of this proposed method.

1 はじめに

FTTHに代表される高速な双方向インフラの普及に伴い、Skype[8]やiChatAV[5]のような音声や動画を利用した多対多コミュニケーションシステムの利用が進んでいる。

ここで、既存のインターネットにおける多対多音声チャットシステムをコミュニケーションの形態に着目して分析する。本論文では、コミュニケーションの形態をコミュニケーション空間（参加者同士が直接コミュニケーションを取ることが可能な空間）における話題に着目し、コミュニケーション空間に話題が1つしか存在しない形態を会議型コミュニケーション(図1(a))、話題が複数存在する形態を「場」共有型コミュニケーションと定義する(図1(b))。この定義に照らし合わせると、既存のリアルタイムコミュニケーションシステムは全て会議型に分類できる。

会議型のシステムは、特定の話題に関して議論するのに適したコミュニケーション形態である。しかし多くの場合同時に発言できるのは一人であり、特定の参加者による場の独占が起こり会話の活性化を阻害する可能性がある。また会話の流れから外れる発言は抑制されるため、話題に興味がない（知らない）参加者はその話題が続く間は発言の機会が減るという特徴がある。こういった発言者の偏りや発言機会の減少は会話の参加者が増えるにつれ顕著になる。

一方で「場」共有型コミュニケーションは、現実世界での立食パーティーやコンパのようにコミュニケーション空間に複数の話題が存在することが可能である。参加者は複数の話題を把握することが可能なので、自由に話題を選択できる。また参加者は新たな話題の生成も可能である。そのため会議型システムと比べ場の独占や発言機会の減少が発生しづらく、会話の発展性や自由度が高いコミュニケーション形態である。

このように「場」共有型コミュニケーションには、会議型コミュニケーションとは違った利点をもつが、現状ではインターネット上での利用は進んでいない。

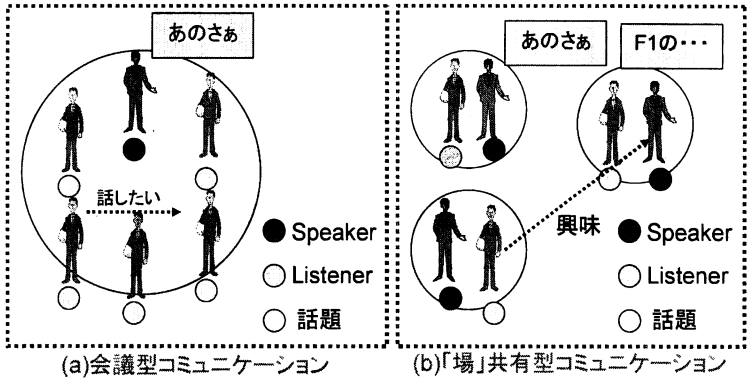

図 1: コミュニケーションの形態

その原因としては、コミュニケーション空間に複数の話題が存在した場合、その識別が難しくなるという問題に起因すると考えられる。

本論文ではこうした背景や問題点を受け、コミュニケーション空間上で複数の話題が並存可能な「場」共有型のリアルタイムコミュニケーションシステムを提案し、実装する。

本論文では、まず「場」共有型コミュニケーションの要件について述べ、関連研究の説明とその問題点の指摘を行う。次に本研究における提案を示し、本システムの設計と実装について説明する。そして評価実験によって提案の有効性を検証する。

2 「場」共有型コミュニケーションの要件

本研究で提案する「場」共有型のコミュニケーションの要件として、コミュニケーション空間内で特定の話題について会話する参加者のグループ（会話グループ）とその会話に着目し、以下の３つを挙げる。

- 会話の並存
  コミュニケーション空間に複数の会話グループが並存できる必要がある。
- 参加会話の聞き取りやすさ
  参加した会話グループの発言は、他の会話グループの発言よりも重要度は高い。そのため非参加会話に比べてより聞き取りやすい必要がある。
- 非参加会話の概要把握
  「場」を共有するためには、会話グループに参加しながら、コミュニケーション空間に存在する他の会話グループの会話内容をある程度把握できる必要がある。これによって参加者は会話に参加しながらより興味のある会話グループを自由に選択でき、会話の自由度が達成される。

3 関連研究

複数の話題が並存可能なコミュニケーション空間を実現する既存研究として、3D サウンドを利用した音声チャットシステムである Voiscape[1, 2] と RAVITAS[3] がある。

3D サウンドは人間の頭部周辺の音響特性を再現することで臨場感のある 3 次元の立体音響を作り出す機能のことである。この機能により、一般的なヘッドホンやスピーカを利用する環境においても、仮想的な音源の位置に応じた音響を実現できる。例えば 3D ゲームの仮想空間において、左前方に存在するアバターの音声が、まるで実際にその位置にいるかのような臨場感をもたらすことが出来る。3D サウンドを実現するライブラリとしては、マルチメディアライブラリである DirectX の一機能である DirectSound[4] や OpenAL[7] がある。

3.1 関連研究の問題点

3D サウンドによって複数の会話が並存する環境を実現できるが、現段階ではデバイスやライブラリの性能上の限界から、現実世界と同等の精度で音源の位置を認識（音源定位）することはできない。そのため複数の会話が並存した場合に会話が聞き取りにくくなるという問題がある。

参加会話の聞き取りやすさに関して、RAVITAS は参加会話の音量をあげ非参加会話の音量を下げるという手法で聞き取りやすさを実現している。それに対し Voiscape はそのような工夫を行っていないため、近くに別の会話グループが存在する場合、参加会話が聞き取りにくくなるという問題がある。

また非参加会話の概要把握に関しては、Voiscape、RAVITAS ともに考慮されていない。そのため非参加会話グループの位置（遠い・他の会話グループと近接している等）や、参加会話による干渉により聞き取れない状況が発生しうる。

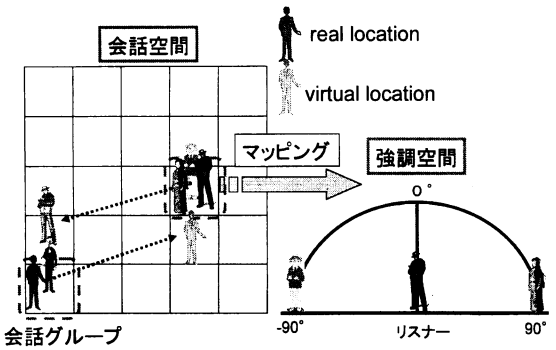

図 2: コミュニケーション空間

4 本研究の提案

2節で述べた要件を満たす「場」共有型のコミュニケーションシステムを実現するためには、参加会話の聞き取りやすさを向上させつつ、同時に非参加会話の概要把握を実現する必要がある。これはRAVITASのように参加会話・非参加会話の音量を調整するだけでは実現できない。

そこでこの要件を満たすために、本研究では会話空間と強調空間から構成される二重構造のコミュニケーション空間を提案する。

次項で提案するコミュニケーション空間について説明する。

4.1 コミュニケーション空間

本研究で提案するコミュニケーション空間は、会話空間と強調空間というそれぞれ独立した2つの空間から構成される。会話空間によって**会話の並存**を、強調空間によって**参加会話の聞き取りやすさ**を、会話空間と強調空間を併用する会話の探索機能によって**非参加会話の概要把握**を実現する。

4.1.1 会話空間

会話空間はシステム全体の参加者同士で共有される空間で、参加者は自由に空間内を移動し任意の会話グループに参加できる。会話空間は2次元平面であり均等な大きさのグリッドに分割されており、同じグリッド上に存在する参加者が会話グループを形成する (図2)。また会話空間における距離や方向を反映した音の臨場感を3Dサウンドによって実現する。参加者は会話空間における実際の位置である real location と、会話の探索に使用する virtual location という2つの位置を持つ。virtual location については後述する。

会話空間は、GUI 上では現実世界に近い3Dの仮想空間として表示し、利用者が直感的に音源の位置を理解出来ることを目指す (図 3(a))。また参加者をアバターで表示し (図 3(b))、発言状況をインジケータで示すことによって、発言者と音源のマッピングを支援する。

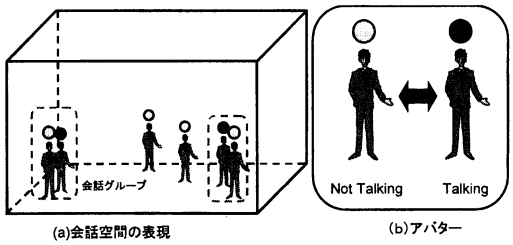

図 3: 会話空間とアバター

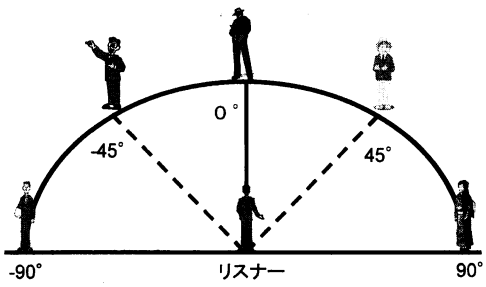

図 4: 強調空間

4.1.2 強調空間

強調空間は、参加した会話グループの発言を強調するために利用される参加者固有の空間である。会話グループに参加すると、会話グループの他のメンバーは強調空間にマッピングされる。強調空間は既存の3Dサウンド環境で音声が聞き取りやすい位置である、前方90°から-90°までの参加者の近傍の空間である。この空間に、発言に応じて会話グループの参加者をマッピングする。

マッピングは一定期間内の発言者数に応じて動的に変更する。発言者が一人の場合は前方に、二人に増えた場合は左右に、さらに発言者が増えた場合は均等に左右の角度を分割した位置に配置していく (図 4)。一定時間発言がなかったり、会話グループから離脱した場合は、その発言者のいた場所を空ける。

会話の強調は音源の移動だけではなく、ユーザインタフェース上での視覚的な強調も同時に行われる。視覚的な強調はユーザインタフェース上の特定の領域を強調画面として利用し、発言に応じて強調空間での位置に応じた場所に参加者の画像を表示することで行う (図 5)。

このように、会話の強調は単純に音量を調節するのではなく、音源の移動と視覚情報による補助により行う。これによって音源定位のしやすい状態を作り上げる。この音源定位によって複数の会話がある状態で会話の聞き取りやすさの向上が期待できる。この手法は音量調節のように排他的に特定の会話を強調する手法とは異なり、あくまで音源定位を実現しているだけなので、特定の会話を強調しつつ別の会話にも注意を向けることが可能である。

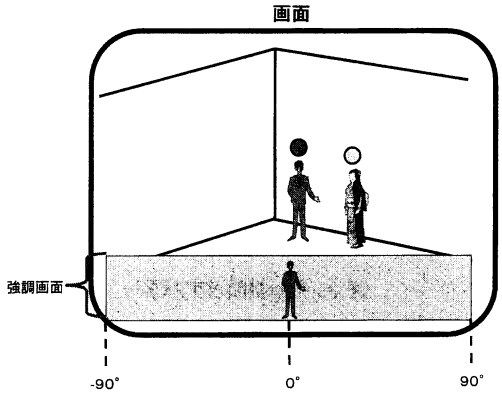

図 5: 強調画面

4.1.3 会話の探索

会話グループに参加し会話の強調機能を利用した状態で、自由に会話空間を移動し、別の会話グループの会話を聞くことが出来る機能を実現する (図 2)。この機能を会話の探索と名付ける。

会話の強調機能を利用しているため、会話空間のどこにいても常に参加した会話グループの発言は聞き取ることが出来る。さらに会話空間に存在する任意の会話グループに接近することで、その会話の概要も把握することが出来る。このように会話空間と強調空間を組み合わせて利用することで、非参加会話の概要把握を可能にする。

この会話の探索時の移動は、virtual location の変更という形で行われる。virtual location の移動は他の参加者に通知されず、他の参加者からは探索者は real location にいるように見え、発言もその位置から聞こえる。

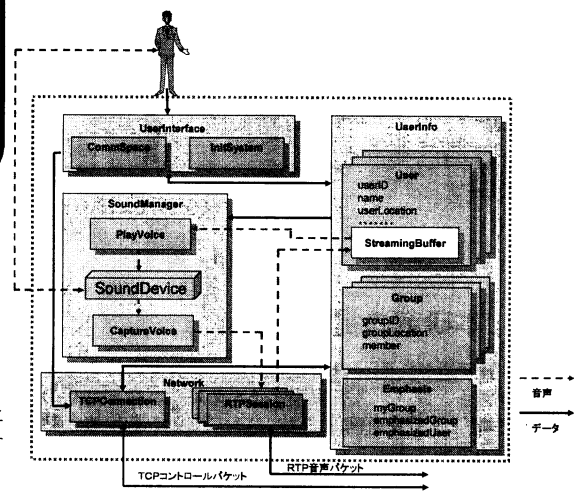

図 6: モジュール構成

5 設計と実装

提案システムの機能に着目し、システムをユーザインタフェース、ユーザ情報管理、音声処理、ネットワークの 4 機能に分割する。この 4 機能について、それぞれを UserInterface、UserInfo、SoundManager、Network の 4 つのモジュールとして設計する (図 6)。以下各モジュールについて概説する。

UserInterface
利用者のユーザインタフェースとなるモジュールである。画面表示と各種操作を実現する

SoundManager
利用者の音声のキャプチャ(CaptureVoice) と、3D サウンドを利用した音声の再生 (PlayVoice) を実現する。

UserInfo
ユーザ名やコミュニケーション空間での位置などの参加者の情報を一元管理するモジュールである。

Network
音声通信を行う音声通信機能 (RTPSession) と、P2P ネットワークを構築しインデックス情報などの参加

図 7: 実行画面

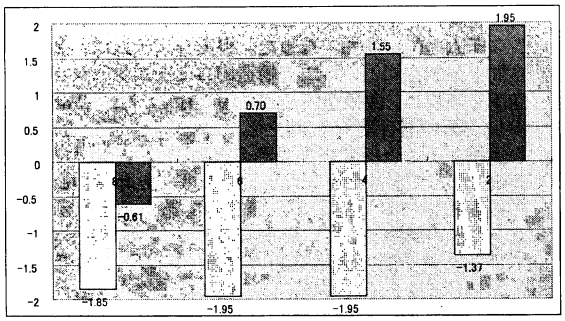

図 8: 参加会話の聞き取りやすさ（アンケート結果）

者に関する情報の交換を行うデータ通信機能 (TCP-Connection) から構成される。

これらのモジュール間でデータや音声の受け渡しが行われ、利用者とのインタラクション、音声のキャプチャ・再生、P2P ネットワークの構築・維持といった処理が並列して実行される。

UserInterface、SoundManger はそれぞれ DirectX の機能である Direct3D と DirectSound、TCPConnection は Winsock[1]、RTPSession は JRTPLIB3.3.0[6] を利用して実装した。

実行画面の例を図 7 に示す。アバターは立方体に任意の画像をテクスチャとしてマッピングしたオブジェクトである。アバターの上方には発言状況を色で示すインジケーターを表示し、音声とアバターのマッピングを補助する。画面下部には強調空間がマッピングされ、会話グループに参加している際は参加者の発言に応じて発言者が表示され、会話の強調が行われる。

6 評価

提案手法の有効性を確認するため、10 人の被験者に対し参加会話の聞き取りやすさと非参加会話の概要把握を確認する評価実験を行った。

評価環境として一辺２０ m の正方形の会話空間を構築し、録音した会話を会話空間に配置した。指定した会話の聞き取りやすさについて 2 点（大変聞き取りやすい）から-2 点（大変聞き取りにくい）までの 5 段階評価でアンケートを行った。同時に発言の概要も記述してもらい、実際の会話内容と一致しているかをチェックした。

6.1 参加会話の聞き取りやすさ

周囲に複数の非参加会話がある場合における参加会話の聞き取りやすさを評価した。非参加会話の数を 8 つから 2 つまで 2 つずつ減らしていき、参加会話の

[1]Windows で TCP/IP の機能を利用したソフトウェアを開発するための API

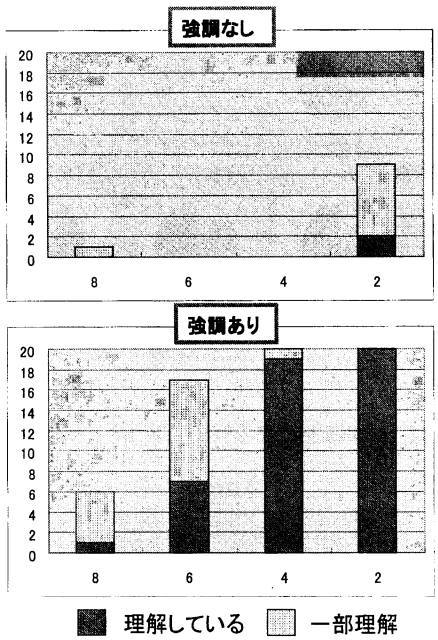

図 9: 参加会話の聞き取りやすさ（把握した概要）

聞き取りやすさに関してアンケートを行った。このアンケートを、会話の強調をした場合としない場合について、各被験者に対してそれぞれ 2 回づつ行った。

評価結果から、会話の強調機能が参加会話の聞き取りやすさに大きく貢献していることが分かる（図 8)。実際に把握した概要に関しても同様の傾向が見られる（図 9)。特に非参加会話の数が少なくなるにつれその差が大きくなっている。これは会話の強調機能を利用しない場合は、少数の会話しかない状態でも参加会話の識別が難しいためだと考えられる。それに対し会話の強調機能を利用した場合は、音源の位置が移動され視覚的に強調されて表示されることで、複数の音源の中から対象となる音源の識別が容易になり、聞き取りやすさが向上したと考えられる。

6.2 非参加会話の概要把握

会話に参加した状態で、非参加会話の概要が把握できるかを評価した。非参加会話は 3 つ用意し、参加した会話グループから近い順にそれぞれ会話 1、会話 2、会話 3 とする。会話の探索機能を利用した場合としなかった場合に対し、各会話の聞き取りやすさについてアンケートを行った。このアンケートを、会話の探索機能を利用した場合としない場合について、各被験者に対してそれぞれ 2 回づつ行った。

会話の探索機能を利用しない場合は、会話グループの距離に応じて聞き取りやすさが下がっている。それに対し会話の探索機能を利用した場合は、会話の距離によらずいずれにおいても聞き取りやすさの点で利用しない場合を大きく上回った（図 10)。把握した概

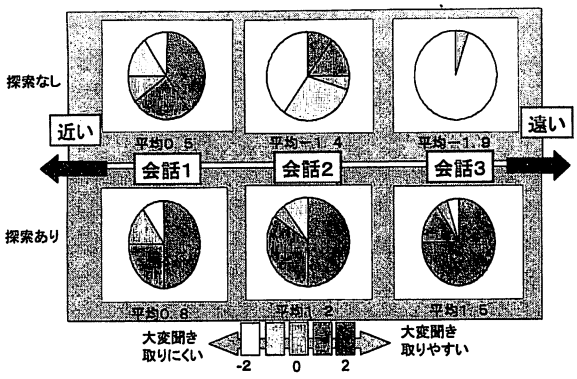

図 10: 非参加会話の概要把握（アンケート）

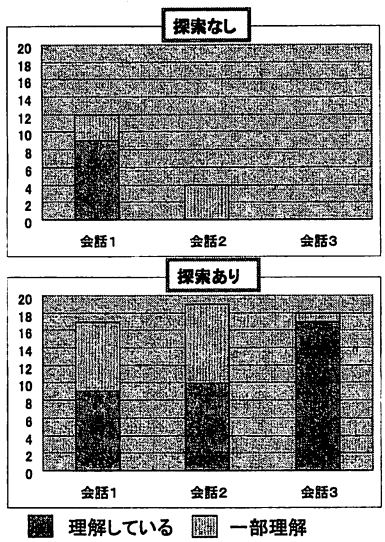

図 11: 非参加会話の概要把握（把握した概要）

要に関しても同様である（図 11）。会話の探索機能を利用した場合、参加者は任意の会話グループに対して聞き取りやすい位置に移動して会話を聞くことが出来るため、このような評価結果となったと考えられる。

この評価結果から、会話の探索機能は非参加会話の概要把握に対し有効であるといえる。一方で聞き取りやすさに関する意見ではないが、強調空間にマッピングされた参加者達の位置が、会話空間上での位置関係と異なることに対する違和感を訴える被験者がいた。実際には独立した空間として扱っているが、2つの空間を同時に視認することが出来る以上、位置関係の整合性を取る必要がある。この点は今後の課題である。

7 まとめ

「場」共有型コミュニケーションの3つの要件（会話の並存・参加会話の聞き取りやすさ・非参加会話の概要把握）を挙げ、要件を満たすコミュニケーションシステムの提案をおこなった。提案システムは、会話空間と強調空間という二重構造のコミュニケーション空間を実現し、会話空間で会話の並存を、強調空間で参加会話の聞き取りやすさを、会話空間と強調空間を連携することで非参加会話の概要把握を実現した。評価によって参加会話の聞き取りやすさと非参加会話の概要把握に対し提案手法が有効であることを示した。今後の課題として、参加者の違和感の原因となっていた強調空間と会話空間の位置関係の不一致を解消する必要がある。強調空間の位置関係にあわせ、動的に会話空間における会話グループに属する参加者の位置を変更する機能を実現すれば問題を解決できると思われる。またより自由度の高い会話を実現するために、会話グループのモデルを見直す必要がある。現状では会話空間にマッピングされた単純な構造であるが、会話の流れを反映したグルーピングが可能なより柔軟なモデルを導入したいと考える。

参考文献

[1] Y. Kanada. Multi-context voice communication controlled by using an auditory virtual space. In *2nd IASTED International Conference on Communication and Computer Networks (CCN 2004)*, Nov 2004.

[2] 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 マルチメディアと仮想環境研究会, Oct 2003.

[3] K. Yasumoto and K. Nahrstedt. Ravitas: Realistic voice chat framework for cooperative virtual spaces. In *IEEE 2005 Int'l. Conf. on Multimedia and Expo (ICME2005)*, July 2005.

[4] DirectX. `http://www.openal.org/`.

[5] iChatAV. `http://www.apple.com/macosx/features/ichat/`.

[6] JRTPLIB. `http://research.edm.luc.ac.be/jori/jrtplib/jrtplib.html`.

[7] OpenAL. `http://www.openal.org/`.

[8] Skype. `http://www.skype.com/`.