

# 時空間状況を考慮した複数人音声チャットシステム

## Multi-Party Voice Chat System Embedded in Spatio-Temporal Situations

中蔵 聡哉\*<sup>1</sup>      角 康之\*<sup>1</sup>      西田 豊明\*<sup>1</sup>  
Toshiya Nakakura      Yasuyuki Sumi      Toyoaki Nishida

\*<sup>1</sup>京都大学 情報学研究科  
Kyoto University Graduate School of Informatics

Abstract: Members in the distance don't initiate a conversation by chance, so they have fewer chances of information sharing than members in the same room have. We thought a always-connected sound system is good for the infrastructure which enable us to initiate conversation in the distance. In this paper, we tries to develop such a system and test it. As a result, some problems came out , however it proved that the system is suitable to the infrastructure and showed many challenging findings.

### 1. はじめに

本研究では、利用者全員を常時接続する音声システムを開発した。このシステムは、遠隔非同期状態においての雰囲気の有と偶発的な会話発生への支援を目的とする。

同室状態であれば、直接会話をしていなくても、聞こえてきた音や視覚情報などから自然と情報が共有される。また、ふと目が合った瞬間に会話が起こったり、部屋の外から興味深い音が聞こえてくればその音についての会話が始まったりする。遠隔非同期な状態において、このような事はまれにしか起こらない。この原因は雰囲気を共有できない事にあると考える。多種多様な通信メディアを備え、情報化社会といわれる現代において、遠隔非同期な状態において雰囲気を共有できていないのは、通信システムが基本的に会話を行う時に接続するというモデルを取っている事が原因であると考えられる。会話を行うときにしか通信回線を接続しないモデルではこれらの情報は全て切り捨てられてしまっているのである。

この情報を拾うため、利用者全員を音的に常時接続し仮想空間上に配置する事で仮想的な大部屋にいる状態を作り出す事を考える。このことにより、他のユーザが実世界で別の人と行っている会話や周辺の音が聞こえてくる事になり、相手が置かれている状況を漠然と知る事ができる。また、各ユーザが聞くべき音を聞けるようにシステムが音量制御を行う事を考える。

以上のことから本研究では、常時接続の音声メディアを用いれば偶発的な会話の発生が支援でき、通信回線を開いてまで伝えないような日常的な出来事が伝わり会話が促進される事で、メンバの興味や現在取り組んでいる事柄など知識共有よりも一段低いレベルでの情報共有を行う環境が構築できると仮定しシステムを実装した。

### 2. 関連研究

回線非接続時の相手の状態の把握に着目した研究に Context Aware Messaging Service[中西 2001] があげられる。この研究は、位置情報やスケジュール情報などからユーザの状況を知る事により電話をかける際の問題を緩和してコミュニケーションを円滑化するものである。この研究は、相手の状況を知ったユーザが、電話をかけるか否かもしくは別の手段で連絡を取る

かという選択肢の中から自分で選択するというものであった。我々はより積極的に会話に利用する事を考える。

仮想空間にユーザを配置して会話場を提供するシステムには、FreeWalk[Nakanishi1996] があげられる。FreeWalk では仮想空間内での距離に応じて音量制御を行っているなど類似点が多々あるが、本システムと最も異なるのはその設計思想であり、FreeWalk があくまで仮想空間内部での会話を支援するものであるのに対し、本システムでは、対面している相手とシステム越しの相手を区別する事なく会話を行う事を考える。

本研究では、最終的には情報空間にある知識で人の会話を強化する事を考えている。そのような研究として、電話対話に参加するパーソナルエージェント [角 2002] があげられる。この研究は電話会話に参加するというものであったが、本研究ではより一般的に、日常会話に参加して情報強化を行う予定である。

### 3. 遠隔非同期状態での会話発生

この節では、常時接続する事でどのように会話発生が促進されるのかをまとめる。まず 3.1 節で本システムで用いるコンテキスト情報について定義する。続いて 3.2 節で空間を越えた会話発生、3.3 節で時間を越えた会話発生について述べ、3.4 節で実現における課題と解決案についてまとめる。

#### 3.1 コンテキストの定義

本研究ではユーザの置かれる状況を推測するため、位置、時間、スケジュール情報をコンテキスト情報として扱う。これらの情報から、例えばミーティングを行っているかどうかは予定と時刻から判断できるし、そのミーティングに興味がある(と思われる)人間は、ミーティングの参加履歴から調べる事ができる。ミーティング中の人と意味空間上で近づける事でミーティングの会話が大きく聞こえ、たとえ遠隔地に出張中であろうとミーティングに参加する事ができる。なお、これらは本論分執筆の段階で実装が完了している情報であり、話題や会話の盛り上がりなど、さらに多くの情報を利用する予定である。

#### 3.2 空間を越えた会話発生

遠隔地では通信回線の接続操作に手間がかかる事と、相手のおかれている状況が分からないためむやみに話しかけると邪魔になってしまう可能性があるため、重要な用件以外は見送ったり、掲示板などの文字ベースのメディアを利用したりしがちである。遠隔地での気軽なおしゃべりを行うのは、通信手段が多数用意されている現代においても難しい。

連絡先: 中蔵 聡哉, 京都大学 情報学研究科, 京都市左京区吉田本町 (工学部 10 号館 214 号室), 075-753-5371, 075-753-4961, nakakura@ii.ist.i.kyoto-u.ac.jp

本システムでは、ユーザ全員がお互いに常時接続する事により、聞こえてくる音声から雰囲気や伝わることと回線接続の手間が存在しないことで、発話までの問題は大きく軽減され、音声の面においては同室状態と同様の状態を提供する事ができる。すなわち、このシステムはユーザ全員を仮想的な大部屋にしていると同様の状態にするものであると言える。

### 3.3 時間を越えた会話発生

過去の面白い出来事を伝えるときや、有用な経験を抽出するときなど、会話中に過去の会話を引用する事はよくある。コミュニケーションを円滑に進める上において、過去の会話を利用する事は非常に有用である。本システム上で交わした会話は、消費してしまわずに話者のコンテキスト情報を付与して保存し、適切なタイミングで自動再生する。再生された会話を聞いて経験を抽出したユーザがそれに対し音声メモを残したり、面白い会話を聞いてそれに対してコメントをしたりする事ができるようになっている。付け加えられたメッセージは、元の会話を行ったメンバに伝えられ、さらにそれに対してメッセージを付け加える事ができる。

例えば、ある学会で興味深いポスター発表を見た場合、その会話を聞かせる事でその研究の概要を知ることができる。その会話を聞いたメンバが疑問点をコメントすれば、発表者に伝わり、質問に答えたり今後の研究の参考にする事ができる。また、発表の会話の後に質問と答えやディスカッションが残るため、会話データはより有益なものとなる。

このようにして、時間とも空間とも切り離された会話が、内容にだけ結びついて成長していく。

### 3.4 実現における課題と解決案

上述のアイデアを単純に実装すると、ユーザが増加するにつれ多数の音が重なってしまって分からなくなる事が考えられる。人間には音を聞き分ける能力が備わっているため、少数であれば問題はないが、多数になると個人差が生じる。しかし、現状の電話会議システムでなされているように、話す順番を指定して会話を行うという方法では自然に会話を行う事はできない。

そこで、まず多数人が話している状態を考える。会話が行われているとき、各会話ごとに話題が存在する。その中にはユーザが全く興味を示さないもの、非常に有用なものが存在する。各ユーザの聞きたいと思うであろう話題を推測し、その話を大きく、ほかの話を小さく聞かせる事ができれば有効であると思われる。似た興味をもつユーザたちを、お互いの声を優先して聞けるようにすれば、ふと口に出した発言が会話に発展する可能性を高める事ができるし、時間の節で述べた過去の興味深い経験が有効に活用できるユーザ同士を抽出し、過去の会話を聞かせる事で会話発生を支援できる。

これらの操作を円滑に行うために、実世界の物理的配置とは無関係に、コンテキストに応じてシステムが自動的にユーザを仮想空間に配置する。この仮想空間を意味空間と呼ぶ。意味空間でも、物理空間と同じように、近くなれば音量を大きく、遠くなれば音量を小さくする。

## 4. 実装

### 4.1 システムの動作

システム構成図を図1に示す。システムは、クライアント端末とサーバからなり、通信はインターネットもしくは有線・無線LANを介して行う。音声の送受信はUDP/IP、その他の情報の送受信は全てTCP/IPで行う。

クライアントは、ヘッドセットのマイクから入力された音声と、ユーザの物理的な位置情報をサーバに送る。音声の形式は

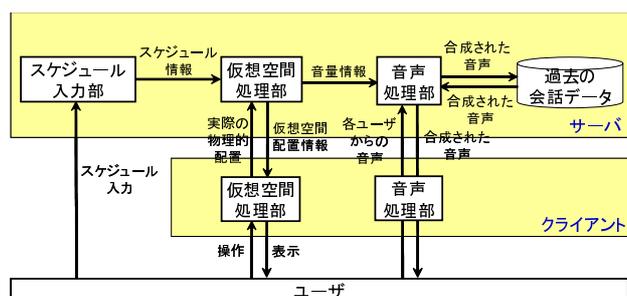


図1: システム構成図

8kHz 16bitPCM であり 20ms 間隔でサンプリングする。送信時にはiLBC のコーデックを用いて圧縮する。位置情報は PlaceEngine[PlaceEngine] を用いて取得する。PlaceEngine とは、あらかじめ位置情報と WIFI の電測情報を登録しておく事で、利用時には WIFI の伝測情報から位置を得る事ができるサービスである。

クライアントから送信された位置情報はサーバの意味空間処理部が受け取り、スケジュール情報入力部の CGI から入力されたスケジュール情報とあわせて利用し、意味空間上でのユーザ配置を計算する。意味空間上での全ユーザの配置情報はクライアントの意味空間処理部へと送信され、クライアント端末のインタフェースで表示される。

クライアントから送信された音声は音声処理部で受信し、意味空間処理部で処理された配置情報を利用して各ユーザの音声を合成する。合成にあたっては、各ユーザからの音声を意味空間上の距離に反比例させて音量を変え線形和をとる。合成された音声はクライアントに送信され、クライアントの音声処理部によって再生される。ユーザには、画面上で近くにいるユーザからの音声が大きく、遠くのユーザからの音声が小さく聞こえる事になる。

ユーザが自由に動き回れるように、クライアントアプリケーションはモバイル PC 上で動作させる。クライアントアプリケーションは、C#および DirectX で実装した。通信は無線 LAN(802.11g)、音声の入出力は耳かけの小型ヘッドセットにより行う。小型ヘッドセットはクライアント端末と Bluetooth(ver1.2) によって接続される。利用中のユーザの写真を図2に示す。



図2: 利用中のユーザ

本システムの操作画面例を

図3に示す。操作や表示が複雑であればあるほど、画面を見る時間が増えてしまい、実世界でのインタラクションを阻害してしまうので、意味空間の状態が一目で分かるシンプルなものに仕上げた。各ユーザの画面表示は同一で意味空間全体を表示している。このインタフェースはモバイル端末のペンデバイスで操作する事を想定しており、クリックとドラッグのみで操作できるように実装した。各ユーザの位置は、ユーザの名前の頭文字の漢字1文字のオブジェクトで表現する。自分以外のユーザのオブジェクトを含む全てのオブジェクトは、ドラッグする事によって位置を変更する事ができる。ユーザにオブジェクトの

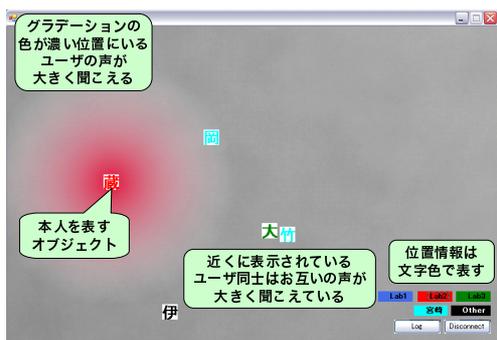


図 3: 「蔵」で表されるユーザの端末の操作画面

移動を許すと、システムが設定した位置関係が壊れる事になるが、手動で自由に操作できる事で話し相手を変更する事や他人を移動させて会話を聞かせる事ができる。システムによる配置との優先関係の最適なバランスについては今後調査する。

## 5. 運用実験

日常的な運用においてこのシステムがどのような影響を及ぼすのかを検証するため、研究室内の3部屋において、自分の机のある部屋に実験参加者を割り振り、日常生活において装置を利用してもらった。時間は2時間、参加者は、当研究室の学生4名、研究員1名、教員1名である。部屋の中を自由に移動できるように通信手段は無線LANを用いた。また、これらの部屋内には実験非参加者も2,3名居た。この実験を2回行ったところ、興味深い現象や意見が多数得られたので以下にまとめる。

### 5.1 実験の結果

以下のような事例と現象が確認された。

	1回目	2回目
他の部屋の会話に参加する	7件	5件
他の部屋の発言を受け会話が発生	4件	2件
相手が聞いているかを確認する	5件	1件
カクテルパーティ現象	4件	1件
自分および他人を移動させる	多数回	多数回

- ・発話のタイミングが重なり、衝突してしまう事がある
- ・画面に表示されている漢字を組み合わせ、ユーザが言葉を作り始める
- ・相手が突然黙るとユーザは画面表示を確認する
- ・全員が一緒に会話を行う傾向がある
- ・遅延や見えない事を利用して会話を盛り上げる
- ・酷い発言を行ったユーザを懲罰として遠ざける

予備段階終了時に行ったインタビューで得られた意見は以下の通りである。

- ・大人数で会話ができるのは楽しかった
- ・他人を動かせるのが楽しい
- ・実際の相手の場所は気にならなかった
- ・遅延が気になった。ギャグが滑ったと思ってしまう
- ・話したくないのに会話に加えられてしまう事がある
- ・声だけでもかなりの情報量がある

また、実験の前後やシステム試用時に、非利用者がユーザを観察した経験から以下のような意見を得られた。

- ・システム利用者が何も無い所に向かって話しているように思える
- ・自分の声が装置越しに誰に届いているのか分からず気持ち悪い
- ・電話とは違い邪魔してはいけないという意識はなかった

### 5.2 考察

2時間の実験を2回行ったが、1回目は11件、2回目7件の部屋をまたがった会話が発生し、これらの会話により情報が共有されている例も多数見られた。全体を通して、ほぼずっといずれかのユーザが話していた。常時接続して全員に聞かせれば会話発生が支援できるという最初の仮説が正しい事の表れであると考えている。

1回目と2回目の会話発生数の差であるが、これは2回目の実験で回数が減少しているのではなく、参加者が初めて見るシステムに興味を示して活発に利用した1回目が多かったためであると考えられる。実際1回目の実験では、会話の内容を理解して参加した例より、聞こえてきた単語に反応して茶々を入れたら会話に発展したというケースが多かった。この点から2回目の7件程度が妥当な数字だと考えている。この結果は実世界においてメンバが同じ部屋にいる状態に発生する回数の平均よりも多いと思われる。実世界においては声が聞こえていても視覚情報から遠くにいる、近くにいうのが明らかであるため、遠くにいる相手には話しかけないというのが原因としてあげられる。このシステムにおいては、声が聞こえてくるメンバ全員を、電話における話し相手と同等の距離感覚で捕らえたからであると考えられる。

部屋 A	部屋 B
A「この環境でレポートするのは凄く難しいなあ…」	B「レポート?何のレポート?」
A「ああ,X先生 X先生.X先生のレポート」	B「ああ…」
A「B(その授業を)とってないやろ?」	B「うん。」
以下省略	

部屋 A	部屋 B
	A「来週は特定領域で…」
B「特定の領域ってどこですか?」	
C「わからないけど多分…」	
以下省略	

その他の興味深い現象として、単純に音量調整のために移動させたわけではなく、ユーザがユーザオブジェクトの漢字を並べ始めた事があげられる。この行動のために、ユーザが他人を移動させた回数が非常に多くなっている。例えば、松、三、角、井、蔵を組み合わせ、松井、角松、三蔵などの文字列を作成した。この現象は、メンバ全員で同じものを見ながらの会話に興味深いということの表れであると考えられる。ログデータにおいても、これらの会話の時間ももっとも長く、また音量も最大であった。

この過程において、遠隔地において、あっち、こっちなどの指示語を用いたコミュニケーションが成立している事が確認された。「あっち」は他人のオブジェクトを自分たちのオブジェクトから遠ざける際に、「こっち」は近づける際に頻りに用いられ

た。不適切な発言を行った者を懲罰として遠ざけるなどの行動も見られた。

また、連続した会話においては問題ないものの、同じタイミングで発言して衝突したり、誰に話しているのか分からなかったりと、目視情報が会話に及ぼす影響が如実に現れた。目視情報に変わる制御法を考える必要がある。逆に、相手に見えていない事を利用してオーバーな表現をして話を盛り上げるなどの現象も見られた。

得られた意見について

おしゃべりをする場合においては、システムは裏方に徹するべきである、勝手に音量を変えられては迷惑である、などの意見が多数得られた。これは会話の進行中にシステムが音量制御を行った場合についての意見であると思われる。話し相手が突然黙るとユーザが表示を確認するという行動が見られた。現状ではユーザが黙ったのか、移動させられて聞こえなくなったのかの区別が付かない。加えてユーザが自発的に移動したのかシステムまたは他のユーザに移動させられたのかの区別もできない。会話進行中である事をシステムがきちんと認識し、会話中であれば制御は行わない、または移動された事を音声的に提示するなどの工夫が必要である。

また、話したくないのに会話に加えられてしまうことがあるという意見が得られた。関連する現象として、利用者に電話がかかって来たときに話しかけられて困るという現象が見られた。状況をより詳しく分析し、会話を行えるかどうかの判断も加えなければならない。

非利用者は、利用者の行動に違和感を抱いたようである。まず、ヘッドセットが小型なため、利用者がどこに向かって話しているのが分からず奇妙な印象を与える。その他にも、システムのマイクがどれ位音を拾うのが分からないために自分の声がどこの誰に聞かれているのが全く分からないという不安感や、目の前の相手が自分と話さずにどこの誰かと話している事に対する疎外感が発生する。システムの利用者同士ではこういった問題は生じなかった。

実際は電話でも同じ問題が発生しているのに、そういった感覚にならないのは、ユーザが電話に慣れ親しんでいる事以上に、装置の外見の持つ印象がもたらす安心感というものがある。ただし、電話とは異なり、会話を邪魔してはいけないという意識は働かないようである。システム非利用者が、システム利用者が同じ部屋にいるから会話を控えるという現象も利用者が遠慮するという現象も確認できなかった。

### 5.3 追加実験

特に制御もタスクも存在しない自由な会話においては、ユーザは全員一箇所に集まり全員で会話を行う傾向があった。これは、最初から全員で話そうとして集まったわけではなく、なされている会話に一人ひとりが加わった結果である。日常的に用いられれば、全員が違うコンテキストを持ち、それによって思い通りに会話を行う事はありえるだろうが、今回の場合は特にそのような欲求を持っていなかったためこの結果になったのであろう。会話の意思がなくなったユーザが、自発的に静かなところに移動する事はあったが、予備期間や本研究で行った実験のような短時間の運用では、分散していくつかの会話が行われている、という状況は見られなかった。

この実験や、その他のユーザが自由にシステムを使用する実験では、全員が1箇所に固まって会話を行い、意味空間の検証ができなかったため、別途強制的にユーザを配置する実験を行った。参加者は本研究室の学生8名と研究者1名である。十分離れている3つの部屋に3人ずつ合計9人のユーザを配置し、全員がシステムを装着した状態で、各チームに分かれて競争して

クロスワードを解く実験を行った。チームは各部屋1人ずつの合計3人を3チーム作った。筆者はオブザーバーとしてシステムにログインしていた。

この過程で、3つの会話が同一のシステム上でなされている事、問題なく意思疎通を行える事が確認された。ただし、ユーザによっては実世界上で直接聞こえてくる声の時々耳に入ってしまい、気が散るといった意見が得られた。自由に使ってもらったときにはこの問題は起こらなかったが、各部屋でなされている会話に外部から誰かが参加するという形になっていたからであろう。また、この実験において、他チームにばれないようにこっそり筆者をチームに移動させ、ヒントを聞いたり、インターネットで調べていいか質問を行うなど、意味空間を制御して排他的に会話を行うという現象がみられた。

## 6. おわりに

常時接続の音声チャットシステムによる偶発的会話発生への支援環境を構築し、システムの開発と運用実験を行った。その結果、実世界における会話発生よりも、このシステムを用いた場合の方がより多くの会話が発生することが明らかになった。これは実世界の大部屋においては遠くにいる相手には通常話しかけないが、このシステムでは視覚情報がないため、声が聞こえてくるメンバ全員を電話における話し相手と同等の距離感で捕らえたからであると考えられる。この他にもさまざまな興味深い現象が観測された。しかし未だ実装上の問題、非利用者の受ける印象など環境として用いる際に多数の問題があることも同時に明らかになった。今後は、システムによる自動配置精度の向上のためにコンテキスト情報を増やすとともに、誰が聞いているのかやどのグループがどんな会話をしているのかを画面表示に頼らず音声的に表示する、Web上の情報を用いて情報強化するなどシステムを改良していきたい。

## 謝辞

システム開発などに関して数々の助言を頂いた久保田秀和氏、VoIP作成にあたりご教示いただいた中西靖明氏、藤川賢治氏に感謝致します。

## 参考文献

- [中西 2001] 中西泰仁, 辻貴孝, 大山実, 箱崎勝也: Context Aware Messaging Service:位置情報とスケジュール情報を用いたコミュニケーションシステムの構築および運用実験, 情報処理学会論文誌, Vol.42, pp.1847-1857 2001.
- [Nakanishi1996] Nakanishi, H., Yoshida, C., Nishimura, T. and Ishida, T.: FreeWalk: Supporting Casual Meetings in a Network, International Conference on Computer Supported Cooperative Work, CSCW96, pp.308-314 1996.
- [角 2002] 角 康之, 出山 敦祥, 間瀬 健二: 電話対話に参加するパーソナルエージェント, エージェント合同ワークショップ (JAWS 2002), 電子情報通信学会, 日本ソフトウェア科学会, 函館, 2002.
- [PlaceEngine] Sony Computer Science Laboratories Inc.: PlaceEngine, <http://www.placeengine.com/>.
- [ProjectSynvie] ProjectSynvie: Project Synvie, <http://synvie.net/>.