

発話が空間コンテキストに埋め込まれることによる 体験共有方法の検証

藤倉 稜^{1,a)} 角 康之^{1,b)}

概要：本稿では、発話が空間コンテキストに埋め込まれることにより、体験共有を促進する方法について述べる。日常生活の中で、我々人間は他人との会話や独り言といった発話を常に行っている。その中には記録される発話もあれば、その場限りの発話も存在する。そこで本研究では、人間が行う発話に着目し、それを HoloLens を使用して空間コンテキストに埋め込んで音声として記録する。それを、空間コンテキストに従って他人に提示することで感動や気づき、知識などを共有する方法を提案する。おわりに、実際に本方法を用いて、空間コンテキストに発話を埋め込んで得られた知見や課題について示す。

1. はじめに

本研究では、発話が空間コンテキストに埋め込まれることにより、体験共有を促進する方法を提案する。

人間が生活するうえで、何か身をもって経験した時それは体験といえる。体験というのはイベントなどの非日常的な体験と、何気ない行動から発生する日常的な体験に分けられると考える。前者を他人に共有する場合、体験の様子を撮影したりメモを取ったり情報を記録しようとする。しかし、後者の日常的な体験は、記録がされず他人に共有されない場合がしばしばある。例えば、美術館にて「この絵は、作者が母親を亡くした時の感情を表現したものなんだって～」などと誰かが聞いたときに面白いだとか、有用地と思われる情報を含んだ発話がされた。この時、発話者や周囲の人が、この発話を何らかの形で残そうという考えに及ばない場合、共有されることがないままその発話は埋もれてしまう。美術館という知的空間にて、このような他人の知見を含んだ発話が共有されないことは、非常に惜しいことであると筆者は考える。人間の創造的問題解決行動や知的生産活動における発想支援システムの構築の分野でも、他人の知識を情報として取り入れることは重要視されている [1]。美術館に限らず、我々の日常生活の中には、共有されていない価値のある発話が多く存在すると考えられる。

そこで、本研究では、発話を自動的に記録し場所に埋め込むことで、それをその場にきた他人に提示する方法を提

案する。発話を提示する手段として、視覚的に発話を提示する方法と聴覚的に発話を提示する方法の2通りが考えられる。今回は、後者を選択した。その理由として、視覚的に発話の内容を提示したとしても、その場を通りかかる人は発話に注目せず、そこにある展示物だったり看板等、当初の目的の対象に気をとられると考えられるからである。対して、聴覚的に提示する場合、近くを通りかかるだけで発話が提示されるため、聞き手が発話に注目していなくても、興味のある内容を含んだ発話があれば聞き入り、情報を集積することができる。しかし、発話を聴覚的に提示する場合、一度に提示することのできる発話の数は限られてしまう。そこで、本方法では空間コンテキストによって発話を分類し、聞き手の空間コンテキストに従って発話を選択し提示する。ここでいう空間コンテキストとは、発話者はどこにいたのかといった位置情報や、何に注目してどこを向いていたか、発話中は移動をしていたかどうか、手などの身体の動きはどうだったのか、どういう発話を聞いてここまで辿り着いたのかなどといった空間的な状況のことをまとめて指す。発話者の空間コンテキストと発話を紐付けることで、発話に属性が与えられる。聞き手には、聞き手の空間コンテキストと類似した属性を持つ発話が提示されることで、聞き手の興味に近い発話が選択されるのではないかと考える。

具体的な手法として、図1に示した動作シナリオのように、1) 発話を空間コンテキストに埋め込む、2) 埋め込まれた発話を近くを通る人に提示する、といった流れで知識の流通を目指す。図1中では、ある展示空間にて、空間コンテキストに埋め込まれた発話が通りかかった人に提示されている。また、提示された発話により新たに発生した発話

¹ 公立はこだて未来大学大学院

^{a)} r-fujikura@sumilab.org

^{b)} sumi@acm.org

も場所に紐付けられ、空間に新たな知識が蓄積している。これらの流れのループにより知識の流通を目指す。本論文では、発話が空間コンテキストに埋め込まれることにより、体験共有を促進する方法の提案と実装について報告する。



図 1 提案方法の動作シナリオ

2. 関連研究

実世界の状況に応じた情報提供インターフェースの研究例として、Nomadic radio[2] や A wearable spatial conferencing space[3] や LiveSphere[4], Augmentable Reality[5] があげられる。これらの研究では、体験の新たな共有方法により、発見や知識の獲得が促されることを目指すという目的はこの研究と同義であり、音声を取っているという点でもそれらを参照されたい。

Matsumura らは、人と街の間に埋め込まれた会話的知

識を流通させるために、自動車に乗って移動する際の車内会話を記録・提示する手法を提案した [6]。Matsumura らは移動の中にあるエンターテインメント性に注目し、自動車を利用して移動する際に生まれる会話の中で深まる知識を会話的知識 (Conversational Knowledge) と呼んだ。この会話的知識、特に車内会話から生み出される知識を、場所に車内会話を紐付けするという手法で流通させるシステムを提案した。本研究では、車内という限られた状況ではなく人が移動する場面全てを対象として、空間コンテキストに発話を埋め込む事でこの会話的知識の流通を目指す。

また、Müller らは特定の場所に立ち止まるとその場で過去に行われた会話が聞こえてきたり、さらに詳しくポスターや展示物などの特定の場所に注目すると、そういったコンテキストを共有した会話に絞り込まれる Ambient Sound Shower というシステムを開発した [7]。本研究はこの研究と目的が同一であるが、システムの実現をするのに別の手法を用いて再実装しているという位置づけである。

3. 発話を共有する方法の提案

本研究では、発話が空間コンテキストに埋め込まれることにより、体験共有を促進する方法の提案をする。図 1 にて示したように会話や独り言といった発話を場所に埋め込み、埋め込まれた発話が近くを通る人に提示されることで、体験時に生じる発話に含まれる感動や気づき、知識の流通を目指す。図 2 では本システムの流れを示している。以下では、システムを実現するための要件について述べる。

3.1 HoloLens による空間コンテキストの取得と発話の録音と再生

会話や独り言といった発話を空間コンテキストに埋め込むには、発話の音声情報を記録すると同時に位置情報や頭部方向といった詳細なコンテキストを記録する必要がある。本システムではこの条件を満たすデバイスとして Microsoft 社が提供する HoloLens*1 を使用する。HoloLens とは、目の前の現実世界と、3D ホログラムの世界を融合させた Mixed Reality (複合現実) を実現させる HMD(Head Mounted Display) 方式の拡張現実ウェアラブルコンピューターである。HoloLens には深度センサーによる空間認識機能が搭載されており、様々なオブジェクトを現実空間に合わせた位置で設置することができる。また、その空間情報を記憶させることも可能である。さらにマイクも搭載しているため、人間の発話を録音しそれと空間情報を紐付けて空間に残すことが可能である。これらの機能を利用することで、体験における会話や独り言などの発話を空間コンテキストに埋め込む方法を実現する。さらに 3.3 節でも述べるが、HoloLens はスピーカーも搭載しており、他人が追

*1 <https://www.microsoft.com/ja-jp/hololens>

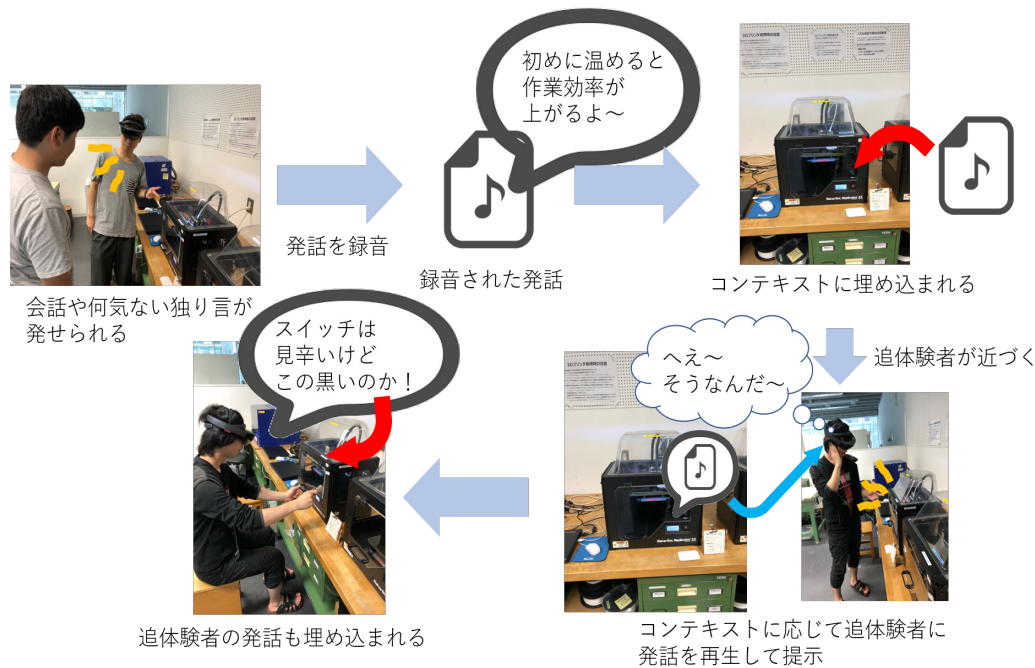


図 2 本方法の流れ

体験する場合 HoloLens をそのまま用いることで一つのデバイスで記録と体験の両方を担うことができる。加えて、HoloLens は、その空間認識機能を使用し、発話者の手など、身体の動きをも認識できる可能性があると考えている。身体の動きは、発話を記録するうえで重要なコンテキストであると考えている。したがって、それを認識できる可能性があるだけで、HoloLens には十分利用する価値がある。

HoloLens を用いた、音声システムの研究として CARA が挙げられる [8]。CARA は、視覚障がい者の移動を支援するために、周囲にある物体の情報を音声で知らせてくれるシステムである。このように、HoloLens を使った聴覚へのアプローチは他でも行われている。

一方、HoloLens は先述した通りウェアラブルコンピューターであるため、無線環境下にて通信を行うことができる。そのため、ある HoloLens を装着して行われた発話を、他の HoloLens に共有し、他人が即座にその発話を聴くということが実現できると考えられる。以上の点より、本方法では HoloLens を利用する。

3.2 発話の切り出しと分類

発話を空間コンテキストに埋め込むとすると、ある程度内容が固まっている発話をひとまとまりとして切り出すことが必要になる。そうすることで、それぞれの発話が行われた状況にできるだけ近いコンテキストに、発話を埋め込むことができる。また、全ての発話が有用な発話であるかどうかは、聞き手によって異なるため発話の分類については議論する必要がある。

発話の分類手法として、岡村らが提案した指差しを手掛

かりにしたシステムがある [9]。一方、HoloLens には指を前に指して、片手の指先 1 本を手前に引くような操作を AirTap と呼び、その動作を検知する機能があらかじめ備わっている。これをトリガーとすることで有用な発話だけを記録できると考える。しかし、1 章で述べた通り、本研究の目的は、誰かが聞いたときに面白いだとか、本来ならば共有されることがないような発話に価値を見出し、有用だと思われる発話を逃さずに記録することである。したがって、発話者の意志に関係なく全ての発話が切り出され記録されるべきであると考えている。また、このようにして切り出された発話は、内容に関連した適切な空間コンテキストに関連付けられる必要がある。

3.3 発話の提示

前節の手法で切り出された発話を追体験者に提示する方法はいくつか考えられる。本節では、HoloLens を使用し、記録した発話を提示する方法について述べる。

まず、記録された発話は発話が行われた場所に紐付けし、記録時と同じ時間幅で再生されるように設定する。聞き手には、本方法によって構築されたアプリケーションがビルドされている HoloLens を装着してもらう。聞き手が、発話が埋め込まれている周辺を歩き回っていると、立ち位置や向いている方向などの空間コンテキストに従って発話が聞こえてくる。また、HoloLens ではステレオサウンドを実現できると考えられるため、どの方向から声が聞こえるといったことも知覚できると考えられる。坊岡らは、各種の建物や施設、商店街や観光地、イベント会場などで、そこに訪れる人が手軽に音声情報を入手できる、AM 微弱電波

を用いた音声案内システムを提案している [10]。この研究では市販の AM ラジオを端末機器として、発信器から出される電波を 2~3m の範囲でキャッチし、音声情報を得る。AM 電波の特性から、発信源に近づけば近づくほど音声は明瞭になるため、視覚障害者にとっては位置把握の手助けにもなることが示されている。このことから、HoloLens のステレオサウンドを利用することで、聞き手を発話により誘導できると考えられる。

さらに、発話の同時提示についても考える必要がある。一度に発話は一つのか、同時に複数の発話を提示するのか、について考えなければならない。本研究の目的である、発話の提示により、新たな発見や知識の獲得が促すことを達成するには、一度に発話をいくつか提示する方が適していると考えられる。複数の発話が混在した場合、必要な情報が聞き取れない可能性があるという懸念は、複数発話の聞き分けにより払拭されることが考えられる。人間は音声について選択的聴取が可能であるため、聞き手は気になる発話を選択的に聴取できるからである。発話を最大同時いくつ提示するかは、現状定めていない。これに関しては、今後議論する必要がある。

また、本方法では、空間コンテキストの類似度によって、提示する発話の選別を行う。例えば、位置情報や頭部方向、手の動きという空間コンテキストに注目した例を紹介する。あるユーザが HoloLens を着用し、本方法を運用しながら工房にて複雑機器の操作について、初学者に説明している場面を想定する。ユーザは複雑機器の前にて、手を所定の位置に動かしながら「初めに温めておくと作業効率が上がるんだよ」という発話を行ったとする。一方、全く別の時間に、別のユーザが本方法を運用しながら工房にて複雑機器を使った操作を行っているとき、先述の発話が複雑機器前にて、スイッチを操作する手の動作に反応して提示される。これにより、発話を聞いたユーザは、複雑機器の操作に関する知見が得ることが期待される。このように空間コンテキストに注目して、提示する発話を選択することは、発話の選別に有効であると考えられる。

4. 発話を共有する方法の実装

本章では実装した方法について説明する。これまで述べてきた機能を備えた、HoloLens 内で実行可能なアプリケーションを開発した。以下では、それぞれの処理について詳しく述べる。開発環境に関して、今回は Unity 2017.2.0f3 と Visual Studio 2017 を使用した。

4.1 発話の検知と録音

発話を埋め込むにはまず HoloLens を用いて発話を録音する必要がある。Unity の AudioSource コンポーネントを用いることで音声の取り扱いが可能となる。今回は、標準で Unity に搭載されている FMOD というサウンドエンジ

ンを採用している。また、マイクからの音声を取得するために Microphone クラスを使用した。発話を記録することは容易だが、3.2 節で述べた通り全ての発話を発話者の意図と関係なく記録する必要がある。そこで、本方法では、録音を常に行い、音量がある閾値を超えた、つまり大きく変化した時点を発話が行われた時点とみなし発話を切り出した。具体的には、マイクの音声から音量を取得して、ある瞬間の音量と 3 秒前の音量を常に比較し、その差が閾値（ここでは 0.03）より大きくなった時点と、小さくなった時点の間の音声を切り出し、発話とした。閾値は筆者の発話を実際に録音して、その音量から適切であろう値を恣意的に定めた。こうすることで、発話者の発話を自動的に切り出す機能を実装した。

4.2 空間コンテキストへの発話の埋め込み

録音した発話を場所に応じて埋め込むために、吹き出しを生成した。それに発話の音声データをアタッチすることで、HoloLens 上では吹き出しから録音した発話が再生されているように見える機能を実装した。こうすることで、発話が終了された時、その時点での位置に、発話が聞こえてくる吹き出しが仮想空間上で自動で生成される。吹き出しを採用した理由については、下記にて説明を行う。

また、その AudioSource コンポーネントの設定を変更し、3 次元的な位置情報を付与した。つまり、埋め込んだ位置と聞き手の位置関係によって、聞こえてくる発話の方向や音量が自動で変化するよう設定した。具体的には、埋め込んだ位置から周囲 8 メートル以内でしか発話は聞こえてこない設定とした。この 8 メートルという数字は、まだ議論する余地があるが、今回は筆者の経験により、発話の内容と関係がある展示物などとの最大の距離を想定し、恣意的に判断した。聞き手が 8 メートル以内に入ると、近付くにつれて発話の音量が 2 次関数的に大きくなる。こうすることで聞き手が発話が埋め込まれている位置の近くを通りかかったときに、発話がざわざわと聞こえてくるようにした。この時、発話の詳細な位置や、発話が埋め込まれている大まかな場所を判断する手助けとして、吹き出しを採用している。今回は、吹き出しのみでしかビジュアルなアプローチは行っていない。また、発話者が止まったまま行った発話ならば、現状の方法で適しているが、移動時に行った発話は一点に留まるのではなく発話者の移動と同じ動きをするのが望ましい。このように、発話を録音し、録音を停止する度に発話が埋め込まれた吹き出しが生成される機能を実装した。

5. システムの動作例とそこから得られた知見

本章では、様々な場面で行った、本方法の評価について述べる。以下では、評価の結果と得られた考察を示す。

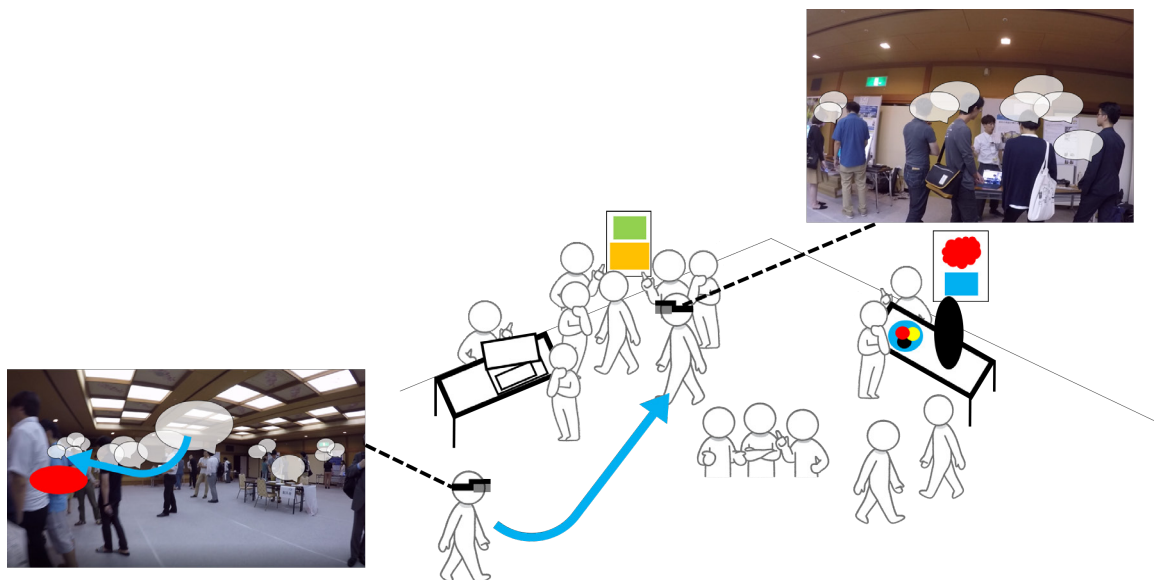


図 3 展示スペースのイメージ図とユーザからの光景

5.1 発話の分類

本方法にて、提示される発話は、聞き手と発話者のコンテキストの類似度によって選択される。これにより、聞き手にできるだけ近い状況にて、過去に行われた発話が提示されるため、聞き手にとって興味のある発話が再生されることが考えられる。

図 3 は、実際に本方法を展示スペースで使用したときの俯瞰図とユーザから見える光景を再現したものである。状況としては、ある国内の学会にて、本方法についてデモ形式で筆者が紹介しているところである。ここでは、本方法を運用していたときに筆者が感じた知見について述べる。ユーザの視点からは展示物や学会の参加者、吹き出しが見えている。この吹き出しは、本方法によって発話が自動的に記録され、そこに埋め込まれていることを示している。つまり、ユーザから吹き出しが見える付近では、過去の発話が聞こえてきている。図 3 左部の写真中では、現在立っている位置から、吹き出しが特定の方向に向かって足跡のように残っている。これは過去の発話者が赤く示された位置につられて、発話を行いながら向かっていったということを示していると考えられる。また、展示を見た他の参加者の発話も埋め込まれているため、吹き出しがあちこちに残っている。これらの発話は、展示に対しての発話者の立ち位置によって、内容の種類と程度が異なると考えられる。実際に発話を聞いてみると、展示に関する発話が残っていたが、展示に近い位置で行われた発話は発表者との会話であったり展示物の細部に関するものであった。逆に、展示から 3 メートルほど離れた位置で行われた発話は、その展示に関するだけでなく、そこから見える他の展示に関することが内容に含まれていた。また、ある位置に留まり続けると、その位置にずっと留まりながらされていた発話も聞こえてきた。

次の例でも、会話の内容は立ち位置によって変化する。図 4 は、イカ型ロボットを観賞しているときのユーザから見える光景である。図左部のように、遠くから見るとイカの形をしているという感想を覚えやすいが、近くで見るとロボットの機構に目が行き、立ち位置によって覚える印象が異なるといったことが考えられる。この印象が発話の内容にも影響すると考えられる。例えば、遠くから見てるときは、「イカの形をした工作かな…なんだろう？」といった発話が行われるが、近くから見てるときは、「とても細かい機構をしている…これは動くロボットなんだ！」といった発話が行われると考えられる。このように、マクロな空間スケールで行われた時とミクロな空間スケールで行われた時とで、発話は種類や内容の細細度が変わってくると感じられる。したがって、聞き手に発話を提示する際には、まずマクロな空間スケールで行われた発話を提示し、次にミクロな空間スケールで行われた発話を提示の方が望ましい。そうすることで、聞き手は、過去の発話の内容を理解しやすくなると考えられる。しかし、現在の方法では、吹き出しの表示は発話の種類に関わらず全て同一のものである。将来的に、これらを移動速度に基づいて自動で識別することで発話を分類することは可能である。例えば、歩きながら行われている発話はマクロな空間スケール、立ち止まりながら行われている発話はミクロな空間スケールといったようにである。ゆえに、聞き手の移動速度と対応づければ、最初の方ではマクロスケールの発話で誘導をして、特定の場所に留まったらミクロスケールの発話の提示を行うということが実現できる。

5.2 埋め込まれている発話による影響

本方法について議論するために、我々は国内の学会にてデモ発表を行い、約 50 人の参加者と約 1 時間半の間意見

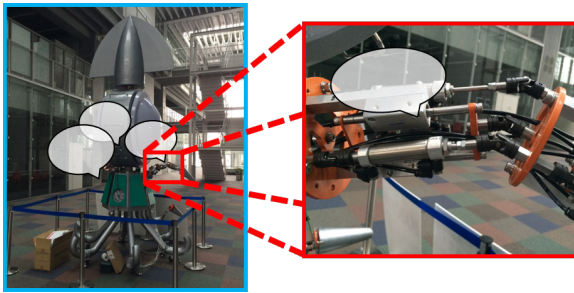


図 4 マクロな空間スケールとミクロな空間スケールとで行われた発話がそれぞれ埋め込まれている例

を交わした。ここでは、その中で得られた印象と発見について説明する。デモの概要としては、本方法について簡単に説明した後、筆者が埋め込んだ発話を自由に移動して聞いてもらった。移動する部屋の大きさは約 150 平方メートルの部屋で、我々を含め 10 件の展示があった。参加者に自由に動いて発話を聞いてもらった後、本方法についての意見を聞いた。まず、体験したほとんどの人が、本方法について面白いと述べた。詳細を続けて聞くと、「他の人が直接話しかけてくるように発話が聞こえてきて、吹き出しがある方へ足が自然と向かっていった」と述べていた。このことから、埋め込まれた発話によって聞き手の行動が誘導されたことがわかった。また、その発話の内容に対して、自ら発話を行いその場に発話を残そうとする人もいた。

一方、今回の方法を開発するうえで、自ら本方法の有用性を感じる瞬間もあった。開発中に、筆者は自分の作業場の周囲にさまざまな発話を埋め込んでることが多々あった。その場では、適当に埋め込んだつもりの発話ばかりであったが、その中に「来週までにこの計画書を出さなきゃ!」という発話があった。後日、アプリケーションの起動したところ、記録された発話が残っていたためその場で再生され、筆者に忘れていた計画書の提出を思い出せた。以上のことから、埋め込まれた発話は自分自身や誰かに影響を与えうると考えることができる。

6. おわりに

本稿では、発話が空間コンテキストに埋め込まれることにより、体験共有を促進する方法について提案した。発話を記録、再生するデバイスとして、発話を空間コンテキストに従って埋め込み可能である HoloLens を使用した。方法としては、HoloLens 着用户の発話を録音し、空間コンテキストに埋め込むことで記録し、それを他人に提示する。本方法を用いて、自分の気づきや知識を含んだ発話を埋め込んだところ、追体験をする人に新たな知識を提供できるのではないかと考察できた。また、本方法を他人に体験してもらったところ、埋め込まれた発話は他人に影響を与えうると考察することができた。

課題として、移動速度に基づいて自動で発話を分類する

機能や更なるコンテキストへの対応の実装が挙げられる。今後の展望として、位置情報や頭部方向だけではなく、発話者の動作を組み合わせることで発話を埋め込むことが挙げられる。具体的には、聞き手が追体験中に発話者と同じ位置にいたとしても、発話者が特定の動作を行ったときに生じた発話は、同じ動作を聞き手がしたときにしか提示されない機能などが考えられる。今後は、これらの機能を実装し、空間コンテキストに埋め込まれた発話が聞き手に提示されることで、新たな発見や知識の獲得が促されるのかを、実際に運用して検証することを目指す。

参考文献

- [1] 國藤進. 発想支援システムの研究開発動向とその課題. 人工知能学会誌, Vol. 8, No. 5, pp. 552–559, sep 1993.
- [2] Nitin Sawhney and Chris Schmandt. Nomadic radio: Speech and audio interaction for contextual messaging in nomadic environments. *ACM Trans. Comput.-Hum. Interact.*, Vol. 7, No. 3, pp. 353–383, September 2000.
- [3] Mark Billinghurst, Jerry Bowskill, Mark Jessop, and Jason Morphet. A wearable spatial conferencing space. In *Wearable Computers, 1998. Digest of Papers. Second International Symposium on*, pp. 76–83. IEEE, 1998.
- [4] Shohei Nagai, Shunichi Kasahara, and Jun Rekimoto. Livesphere: Sharing the surrounding visual environment for immersive experience in remote collaboration. In *Proceedings of the Ninth International Conference on Tangible, Embedded, and Embodied Interaction*, TEI '15, pp. 113–116, New York, NY, USA, 2015. ACM.
- [5] J. Rekimoto, Y. Ayatsuka, and K. Hayashi. Augmentable reality: situated communication through physical and digital spaces. In *Digest of Papers. Second International Symposium on Wearable Computers (Cat. No.98EX215)*, pp. 68–75, Oct 1998.
- [6] Kohei Matsumura and Yasuyuki Sumi. What are you talking about while driving?: An analysis of in-car conversations aimed at conversation sharing. In *Proceedings of the 6th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, AutomotiveUI '14, pp. 34:1–34:8, New York, NY, USA, 2014. ACM.
- [7] Christof Müller, Yasuyuki Sumi, Kenji Mase, and Megumu Tsuchikawa. Experience sharing by retrieving captured conversations using non-verbal features. In *The First ACM Workshop on Continuous Archival and Retrieval of Personal Experiences (CARPE 2004)*, pp. 93–98, 2004.
- [8] Yang Liu, Noelle RB Stiles, and Markus Meister. Augmented reality powers a cognitive assistant for the blind. *eLife*, Vol. 7, p. e37841, nov 2018.
- [9] 岡村剛, 久保田秀和, 角康之, 西田豊明, 塚原裕史, 岩崎弘利. 車内会話の量子化と再利用. 情報処理学会論文誌, Vol. 48, No. 12, pp. 3893–3906, dec 2007.
- [10] 坊岡正之, 相良二郎, 赤澤康史. 微弱電波を用いた音声案内システムの開発. 第 11 回リハ工学カンファレンス講演論文集, pp. 237–238, 1996.