

ウェアラブル環境における音声記憶共有システムの実現

中村 聡史
情報通信研究機構

田中 克己
京都大学大学院 情報学研究科

本研究では、ウェアラブル環境向けの情報共有システムとして、記憶を音声として空間に残し、共有可能とするシステムを実現する。音声を利用することで、ユーザは手軽に記憶を空間に残し、意識することなく記憶を聴取することが可能となる。実現において、音の特性や時間、ユーザの移動などを考慮し、実空間に配置する音声記述言語を提案する。また、空間に配置された音声の問合せ方法についても考慮する。

On Realization of the Sound Memory Sharing System in Wearable Computing Environment

Satoshi NAKAMURA
National Institute of Information
and Communications Technology

Kastumi TANAKA
Graduate School of Informatics,
Kyoto University

In this work, we design and implement the sound information system for wearable computing environment. In our system, users can store their memories as sound contents in real world only to speak and hear/listen sound memories from real space. In order to realize our system, we describe about the specify of sound such as time, location, range, direction and etc. In addition, we define the markup language of sound memories in order to locate them. Additionally, we describe about the method of query for sounds contents.

1 はじめに

ウェアラブルコンピュータの登場 [3] により、人々が空間に記憶を残し、他人とコミュニケーションを図ることが容易になった。Augment-able Reality [14] は、実世界を拡張可能にするシステムである。透過型の HMD (ヘッドマウントディスプレイ) やヘッドホン、マイクなどを装着したユーザが、場所や事物などのコンテキストにテキストメモや音声メモなどの情報を貼り付けることで間接的に情報を伝達する。空気ペン [15] は、透過型の HMD を装着したユーザが、空中で描画ペンデバイスを動かすことで、その空間に手書きのメモを残し、同じく透過型の HMD を装着したユーザとコミュニケーションをとることを可能とするものである。

一方、HMD を使用した情報閲覧システムでは、ユーザは画面に注目しなければ閲覧が困難であることが多く、一般に他の作業との並行性が低いという問題がある。例えば、HMD を利用することで周辺のお店情報を取得することは可能だが、歩きながら画面を見るのは難しく、人や障害物にぶつかるなど

の危険も伴う。また、車の運転中に、目的地までの道順や近隣の地図を HMD に表示することは可能だが、運転をしながら画面を見るのは危険である。以上のことから、HMD は限定的な利用には適しているものの、日々の生活において自然に利用するには適していないと考えられる。また、ユーザは実生活において様々なことに注意しながら様々な作業を行っているため、HMD などを利用して視覚により能動的にコンテンツを閲覧するよりは、聴覚により受動的にコンテンツを聴取するほうが好ましいと考えられる。

そこで本研究では、ウェアラブルコンピューティング環境における自然な記憶の共有を可能とするため、音声のみを用いて記憶共有を図る、音声情報システムを実現する。実現するシステムでは、ユーザはマイクなどを利用して空間に音声情報を記憶する。音声情報は、ユーザが装着しているウェアラブルコンピュータや、周囲に偏在するユビキタスコンピュータに蓄積される。ユーザは、ヘッドホンを利用して空間を移動することにより、空間に記憶 (蓄

積)された音声情報を聞くことができる。この仕組みにより、空間に簡単に記憶を残し、相互の記憶を共有するコミュニケーションシステムが実現される。本研究では、自然なコンテンツ記憶、閲覧を可能とするため、ユーザの歩行スタイルなどの自然なコンテキストを利用する。また、空間性を利用したコンテンツの記憶および聴取を可能とするため、コンテンツ記述言語およびコンテンツ問合せシステムを実現する。

本稿では、まず2章で提案するシステムについて述べ、3章でシステムの設計および実装について述べる。4章でアプリケーション例を示し、5章で考察を行う。6章で関連研究について述べた後、7章で本稿のまとめを行い、今後の課題について述べる。

2 音声記憶共有システム

聞こえた音をすべて紙に書き出しなさい。時間は、二、三分でいい。聞こえた音のリストをつくろう [12].

サウンドスケープ(音の風景)を提唱したカナダの作曲家 R. Murray Schafer による、サウンドスケープ体得に向けた最初の課題である。世界に耳を傾けると、小鳥のさえずり、水の流れる音、風にそよぐ木々の音、ラジオから流れるDJの話し声、子どもたちが遊ぶ音、お寺の鐘の音、駅での電車の案内などなど、多くの音を聴き取ることができる。こうした音は人の音、機械の音、自然の音に分類することが可能であり、我々は目で眺めずとも、聴こえる音から季節、時間、周囲の環境、今起きていること、これから起ころうとしていることなど、多くの情報を意識的、無意識的に得ることが可能である。

一方、過ぎ去ってしまった音を再び聴くことはできない。その場で話されていた会話を後になって聞くことはできないし、祭りの次の日に祭りの音を楽しむことはできない。秋にうぐいすの、冬にセミの鳴き声を聞く事はできない。スピーカーを利用して、音をその空間にとどめることは可能だが、流される音によっては、その場の雰囲気や台無しになってしまう。例えば、古い町並みの景観とそぐわない商店の広告や、雄大な景勝地にそぐわない不必要な音楽がBGMとしてスピーカーから流れることが問題であることは容易に想像がつく。また、駅構内な

どでのスピーカーなどを用いた画一的な放送を鬱陶しく思う人も多く、場合によっては騒音と認識されている [16]。これは、聴覚が視覚に比べ受動的であり、選択できないことに関係がある。

本研究で提案するシステムでは、空間に手軽に音声情報を記憶として残し、その場その場で、自分にあつた音声情報を聴くことを可能とする。音には位置や方向、時間に加え、空間的な移動や広がり、発声した人の情報などが存在する。本研究では、こうした各種の情報を音声と一緒に空間に記憶し、聞き手が音声情報を聴くときに利用する。あなただけという個人的な情報サービスの実現により、空間の雰囲気や損なうなどの問題を回避できると考えられる。

2.1 音声コンテンツ

音声コンテンツは録音するだけで作成可能であり、何気なく聞く事ができ、他の作業との並行性が高いという特徴がある。また、注意を喚起するためのものとして使えるうえ、複数のコンテンツを重ね合わせて提示し、その中から意図する情報のみを聞き取るということも可能である [1]。さらに、音声コンテンツは場を演出することが可能であり、明るい、暗い、軽やか、重い、暖かい、寒いなどさまざまな印象を与える。例えば、人の声にも、感情やリズム、強弱、声質、イントネーションなど様々な特徴がある。本研究で実現する記憶共有システムでは、ユーザの音声をコンテンツとしてそのまま利用する。これにより、コンテンツ内にコンテンツ作成者のイントネーションや声質、感情などの特徴を内包することができる。音声コンテンツを空間に記憶する場合、音声録音者の位置情報や注目している方向などを合せて保存する。この空間情報を利用し、提示の際に複数のコンテンツを空間的に配置することで、コンテンツの方向や、空間的な広がりなどを提示できると考えられる。ここで、ユーザが移動しながら音声録音を行う場合、その録音された音声には動きがあるといえる。音声の動きをコンテンツ情報として保持しておく、録音された町並み案内の動きに合わせて歩行することにより、町並みの散歩を楽しむことができるようになる。音の広がりや、音の空間性を利用することで、空間の雰囲気を作り上げることも可能になる。なお、本システムは、ヘッドホンで音声を提示することができるため、音声を

空間にとどめる手間は少ない。

上記にあげるコンテンツ記憶のための仕組みを実現するため、音声コンテンツは一意のIDをもつものとし、音声データとセットで各種の特徴を保持するコンテンツ記述言語を用意する。

2.2 コンテキスト

センサ技術の進歩により、位置情報だけでなく、ユーザの体調などの生体情報（体温、脈拍、発汗量など）、天気や温度、湿度など様々な情報を取得できるセンサが登場しており、ユーザや環境について細かなコンテキスト取得が可能となってきている。こうしたセンサで取得可能なコンテキストに加え、ユーザの年齢や性別、ユーザの明示的な要求（本が欲しい、ご飯を食べたいなど）や、趣味、嗜好といった個人情報などを利用した状況依存サービスについて研究が盛んに行われている。

音声記憶共有環境では、音声の記憶（音声コンテンツ）がユビキタスコンテンツとして空間に偏在化されている。ユーザのコンピュータは、周囲に問合せを行い、必要な音声コンテンツを取得して適切な音声情報を再生する。ここで、周囲に偏在するコンテンツの中から利用者に適したコンテンツを提示する場合、ユーザとコンテンツとの距離だけを利用してコンテンツ取得を行うと、無駄なコンテンツが多数取得されてしまうという問題がある。また、ユーザの向いている方向を利用しても大差はない。

本研究では、ユーザが自然な形でコンテンツを空間に記憶し、聴取可能とするため、ユーザが普通の生活において何気なく行っている歩行および首振り動作を利用する。例えば、歩行速度を緩めたり立ち止まったりすることや、道を蛇行して歩くことなどは周囲に興味を示していると考えられるし、歩行速度を速めることは周囲へあまり興味をもっていないと考えられる。来た道に戻るの、これまで歩いてきた道の方に興味があることを示していると考えられるし、四つ角を曲がるのは曲がる先に興味を示しているといえる。また、キョロキョロと首を回して周囲を見回しているときには、その近辺で何かしら探し物をしていると考えられるし、振り返るといふ動作は後方へ興味を示しているといえる。こうしたユーザの位置情報に比べ上位にあたるコンテキスト情報を利用することで、自然な形でコンテンツ

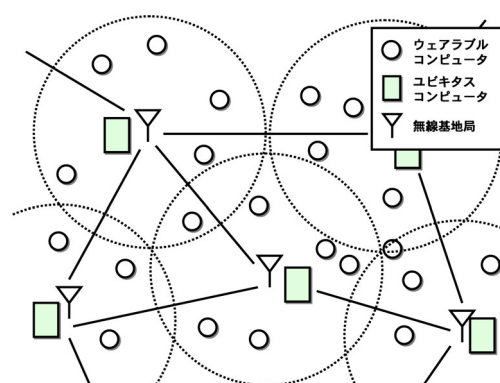


図 1: 想定環境

聴取が可能になると考えられる。また、音声録音を行う場合も、録音中の歩行スタイルや首振りを利用することで、どこら辺に注目して録音を行ったかということ記録することができると考えられる。

他のコンテキストとしては、時間や季節を利用し、時刻や季節にあったコンテンツ提示と、ユーザの明示的な要求に基づくコンテンツを絞り込みを行う。なお、体調情報や天候情報などを利用してコンテンツの記憶および閲覧を可能にすると情報を絞り込むことができると考えられるが、あまりに複雑になりすぎるため、本研究では簡単のためこうしたコンテキストについては考慮しない。

3 設計と実装

3.1 想定環境

想定する環境では、大容量の記憶領域をもつユビキタスコンピュータが実空間に偏在し、有線ネットワークまたは無線ネットワークを利用して相互に接続される、ユビキタスネットワークを構成している。ユーザが装着するウェアラブルコンピュータは無線ネットワーク用のアダプタをもち、空間に偏在する無線ネットワークの基地局を利用してユビキタスネットワークに接続する。本研究では簡単のため、無線ネットワークでカバーできる範囲に、無線ネットワークの基地局と、IPアドレスの管理およびコンテンツの記憶を行うユビキタスコンピュータをセットで設置する。また、隣接する地域のユビキタスコンピュータ同士は有線のネットワークで接続されており、それぞれ隣接するコンピュータのアドレ



図 2: システム構成

スは把握しているものとする。無線ネットワークの範囲に入ったコンピュータは、無線ネットワークの基地局に問合せを行い、アドレスを取得する。ウェアラブルコンピュータは自分のアドレスおよび位置情報がある一定期間ごとに、その地域サーバとして動作するユビキタスコンピュータへと通知する。ユビキタスコンピュータは登録された情報をもとに、近隣のコンピュータリストを管理する(図 1)。

この仕組みにより、無線ネットワークに属しているピアは、ユビキタスコンピュータを介することで近隣のコンピュータに対する通信が可能となる。このネットワークを利用して、コンテンツの蓄積、取得、問合せなどを行う。

3.2 システム構成

本システムは図 2 のような環境で構成される。

ユーザは、ヘッドセット(ヘッドホン+マイク)をつけ、音声の録音および聴取を行う。また、位置や方向を取得するため、ユーザは GPS およびモーションセンサを装着する。さらに、音声録音期間を設定するため、ON/OFF の情報を取得するボタンデバイスを利用する。それぞれの機器は、ユーザが装着しているウェアラブルコンピュータに接続されており、ウェアラブルコンピュータは各種の情報を収集および提示する。

ボタンを押すと録音開始となり、ボタンを押すのをやめると録音終了となる。録音する際の WAVE ファイルのフォーマットは、モノラルのサンプリング周波数 22050Hz で、量子化ビット数は 16 とした。録音された音声ファイルは、後述する音声コンテン

ツの記述言語とセットでユーザのウェアラブルコンピュータおよび近隣のユビキタスコンピュータに蓄積される。

一方、システムはユーザの位置の移動に応じて、先述のネットワークを利用し、動的に音声コンテンツの取得問合せを行う。取得した音声コンテンツを、記述言語をもとに仮想的に空間に配置し、ヘッドホンおよび仮想的な 3 次元音響を利用して空間的に提示する。ユーザはこの仮想的な 3 次元音響(Microsoft 社の DirectSound3D を使用)で提示される音の集合から、情報を得ることが可能となる。

3.3 コンテキスト

ユーザのコンテキストとして、静的なユーザの各種情報(ID, 明示的な要求, 年齢, 性別)に加え、先述のユーザの行動に関するコンテキストを利用する。ユーザの位置や歩行状態を屋内でも取得可能とするため、GPS のかわりに歩行者用自律航法装置である Point Research 社の DRM (Dead Reckoning Module) を腰に装着して取得する。また、NEC TOKIN 社の MDP-A3U9S を用いて、頭部のトラッキングを行う。各センサ情報は 1 秒ごとに取得し、コンテキストとして利用する。ユーザの移動速度がこれまでの平均速度より一定以上ゆっくりになる(または停止する)と周囲に注目しているとし、移動速度が速くなると周囲に興味を示していないと判断する。一方、ユーザが 90 度方向を変えたときには曲がった先に興味を示したことを検出し、180 度逆向きに方向転換したときは、後方に対して興味を示したことを検出する。

腰と頭部の方向に変化が生じたときは、その角度差から左右を向いている、後方を向いているなどを判断する。このとき、ある一定時間内で左右に首を振ったときにはキョロキョロしているものとして周囲に興味を示していることを検出し、首振りがない場合はそちらの方向へ興味を示していることを検出する。同様に、後方を向いているときにも後方へ興味を示していることを検出する。なお、この向いている時間は興味の度合いに比例するものとし、データとして利用する。

これらの情報をコンテキストとして利用し、問合せを行う。

```

<Sound>
<SoundID>*** ファイル名に対応 ***</SoundID>
<AuthorID>*** ユーザごとに一意のID ***</AuthorID>
<Sex>Male</Sex>
<Age>28</Age>
<Activity>Valuation</Activity>
<Date>2004/10/12 20:13:15</Date>
<Location>134.3321, 34.5325, 12.5</Location>
<Direction>0, 1</Direction>
<LocationArray> *** 1秒毎のログ *** </LocationArray>
<DirectionArray> *** 1秒毎のログ *** </DirectionArray>
<Range>10</Range>
<Duration>10.5</Duration>
<Lifetime>1 month</Lifetime>
</Sound>

```

図 3: コンテンツ記述例

3.4 記述言語

今回、システム実現において最低限必要なコンテンツの要素として、下記にあげるものを用意した。Location-Array および Direction-Array は、音声録音における1秒ごとの位置および方向のログである。この情報を利用することで、音声の動きを取得することができる。なお、AuthorID はユーザごとに与えられる一意の ID であり、Duration は音声の録音時間、Range は音声の広がり、Lifetime は音声コンテンツの生存時間を表している。

- SoundID
- AuthorID
- Location (緯度, 経度, 高度)
- Direction
- Location-Array
- Direction-Array
- Date
- User Information (性別, 年齢)
- Range
- Duration
- Activity
- Lifetime

コンテンツの記述自体は、図3のようなXMLライクな言語仕様をとる。

3.5 問合せ

ユーザがコンテンツが記憶されている場所を歩くときに該当するコンテンツの再生を行う必要が

ある。そこで、音声コンテンツは随時先読みして、ローカルキャッシュとして取得しておく。さらに、ユーザの動作に合わせてローカルで計算し、即座に再生できるようにする。つまり、二段階のコンテンツ問合せを行うことになる。

一段階目の問合せでは、ユーザのコンピュータは進行方向からある一定距離までの範囲を指定し、その範囲をカバーしているユビキタスコンピュータへと問合せを送る。このとき、各種のコンテキスト情報も合わせて送付する。ユビキタスコンピュータは問合せを元にローカルのデータベースおよび周辺のウェアラブルコンピュータへと問合せを行い、検索結果を問合せ元に返信する。問合せ元のコンピュータは問合せ結果を利用して、所持していないコンテンツの取得要求を行い、ローカルディスクへと蓄積する。この問合せおよびコンテンツのやりとりにはバケツリレー式のP2Pネットワークを利用する。二段階目の問合せでは、ユーザのコンテキストの変化に伴い、ローカルディスクから適合度の高いコンテンツを複数選び出し、再生する。ここで、適合度はユーザの進行方向の一致度、ユーザの向いている方向の一致度、時間の一致度、ユーザ指定のコンテキストとの一致度の和からなる。この仕組みにより、必要とするコンテンツを取得し、情報を聴取することができる。

4 アプリケーション

実現したシステムを利用することで、空間に道案内情報を残したり、空間に店舗などの評判情報を蓄積したり、人ごとの趣味嗜好に応じた空間のサウンドスケープを実現したり、空間を把握したりすることが可能になる。

例えば、ある建造物への道案内を作成する場合、コンテンツ作成者は道の分岐が近づくと「次は右へ」や「次は左へ」「ここは直進」などのように音声情報を空間に残すだけで、道案内サービスが作成される。その建造物へと行きたい人は、その道案内情報にアクセスし、歩きながら聴こえてくる音声案内を利用することで、道に迷うことなく目的の場所へ辿り着くことが可能になる。観光地などでは、看板などで画一的に道案内を行っているが、情報不足であることが多い。景観を損なうなどの問題がある。本システムを利用することで看板を設置する

必要がなくなるうえ、「〇×というお店を左へ」「あとは道なりに」などのようにより詳しい情報を残すことができるという特徴がある。また、いろんな人が道案内を作成することができるため、自分に合ったコンテンツを選択できるという特徴がある。

また、飲食店の評価を空間に蓄積することが可能となる。その飲食店で食事をした人が、その店の前で「おいしい」「まずい」「〇×がオススメ」「△×はイマイチ」などの評価をする。その音声は空間に蓄積され、次にその店の前に来た人は、空間に残された声情報からこの店の評判を知り、判断材料にすることができる。これは、空間に即したクチコミ情報といえるだろう。

同じ音であっても人によって受ける印象が異なる[17]ため、実世界でのサウンドスケープは難しい。折角考えて配置した音も、人によっては雰囲気損なうものとして悪い印象をもつことがある。また、騒音と感じる人もいるだろう。本システムを利用すると、その場の雰囲気だけを楽しみたい人は音を聞かないように設定し、聞きたい人は自分の好みに合わせて聞く事ができる。つまり、サウンドスケープを個人化することができるため、こうした問題がなくなると考えられる。

本システムを利用することで、空間の把握が容易になると考えられる。人々は、目の届く範囲、音の聴こえる範囲である程度の空間把握ができるが、遠く離れたところの音や映像を見ることはできない。空間に音が記憶されていると、その音声を統合して、ある一定の空間の音として再構成することができる。空間の詳細を切り替えると、単一の音として聴こえていたものは、店舗の音となり、通りの音となり、街の音になる。これを利用することで、どちらの方角にどういった町並みがあるのかなどを把握できる仕組みを実現できると考えられる。

5 考察

実現したシステムを簡易的に利用し、コンテンツの記憶および聴取を行った。利用における感想としては、何気ない歩行でコンテンツの存在を知ることが可能であり、気づかせてくれる楽しいものであるということであった。また、それほど歩く速度が速くない場合は、大体空間を指定したコンテンツ聴取が可能であり、逆に速度が速い場合は（自転車に

乗っているなど）音が聞こえてくるタイミングが遅いなどの問題があった。これについては、ユーザの移動速度に合わせて、コンテンツの先読み範囲を拡大し、取得する仕組みが必要になると考えられる。一方、遠くの音まで聞こえすぎてそのコンテンツがどこに記憶されているかわかりにくいことがあった。この点については、距離による減衰を十分に考慮し、音声コンテンツの聞こえる範囲を絞っていくことが考えられる。また、音の数が増えすぎると聴取が困難になり、鬱陶しく感じるがあった。これは、同じ人が録音したコンテンツが多数存在したということも原因のひとつであるが、聴取中に特定のコンテンツに対してフォーカスを当てるという仕組みがないというのも理由と考えられる。今後は、複数音声の中から注目したい音声コンテンツに対してフォーカスを当てることを可能とするような仕組みを実現する予定である。この音声コンテンツのフォーカシングについては、提示型インタフェース[10]や Dynamic SoundScape[2]などを参考にする予定である。

実現したシステムでは、ユーザが音声コンテンツの録音を行った位置と、実際に音声コンテンツを記憶する位置の距離関係についてはなんら考慮しておらず、音声コンテンツをユーザが録音した場所の周辺に保存される。そのため、街などを歩きながら近くの建物に記憶を残し、記憶された音声コンテンツを聴取することには適しているが、遠方の山や建物などに直接記憶を残すことはできない。つまり、ある山を様々な方向から見ている人同士で同じ音声コンテンツを共有することはできない。しかし、同じ山でも見る距離や見る方向によって印象が異なり、シチュエーションに応じてコンテンツを共有することが好ましいと予想されるため、距離を指定したコンテンツ記憶の必要性は低いと考えられる。

今回実現した記述言語は、コンテンツ記述において十分であるとは言いがたい。例えば、今回実現したシステムでは、空間に偏在化されたコンテンツ同士にリンクなどの構造化情報がない。ユーザ ID や、コンテンツの位置的な接続などを利用し、聴取側である程度の構造化が可能であるが、録音側で明示的な構造化を行うことができない。この音声コンテンツ間のリンク構造については、SoundWeb[11]で用いられている構造化情報と副音声によるリンク部の提示を利用することが考えられる。ここで、ウェ

ブコンテンツとは異なり、ユビキタスコンテンツである音声コンテンツは地理的空間的性をもつため、ユーザの位置や移動に合わせて動的にリンクを変化させるような仕組みが必要になると考えられる。

実現したシステムでは、環境のコンテキストを十分に活用できていない。例えば、晴れた日の午後と、雨の日の午後では聞こえるコンテンツが違うことが好ましいだろうし、暖かい日と寒い日では聞こえるコンテンツが違うことが好ましいだろう。こうした点を考慮し、環境コンテキストを取得するシステムをユビキタスコンピュータなどに接続し、その情報を周辺区域で共有することで、音声コンテンツが録音された時の状態をより詳細に記憶する。問合せ時にもそうした情報を利用することで、細やかなコンテンツ問合せが可能になると考えられる。また、環境のコンテキストだけでなく、ユーザの体調などの生体情報をコンテキストとして利用することも考えられる。この点については今後の課題である。

今回実現したシステムでは、音声コンテンツに各種のユーザ情報が付加されてしまうため、プライバシーを保護できないという問題がある。また、今回実現したシステムでは、音声コンテンツに対する聴取対象とするユーザを指定する仕組みが無いため、誰でもコンテンツを聴くことができってしまうという問題がある。さらに、コンテンツを削除する仕組みが無いため、一旦作成してしまったコンテンツは、期限がきて自動的に削除されるまでは残存してしまう。今後は、こうしたプライバシー保護や、コンテンツに対するアクセス権限の設定、コンテンツの削除システムなどを実現する予定である。一方、現状では空間に対してコンテンツは自由に記憶可能で、聴取も可能となっている。例えば、店内ではコンテンツの記憶はできないようにする、社内コンテンツを社外に持ち出すことを禁止するなど、コンテンツの空間的なアクセス権限が必要になるシチュエーションは多く考えられる。そこで、今後はユビキタスコンテンツの空間アクセス制御についても考えていく予定である。

6 関連研究

音声を用いた情報閲覧システムに関して、さまざまな研究がなされている。

音声入出力を用いて情報閲覧を行うシステムとし

て、Nomadic Radio[6]やAVoN calling[4]、SpeechWear[5, 7]がある。Nomadic Radioは、音声入出力によるテキストメール・ボイスメールの閲覧・作成を可能にするシステムで、ウェアラブルコンピューティング環境における利用に適している。AVoN callingはXMLで記述されたデータを音声入力で操作するシステムである。SpeechWearは、アメリカ海兵隊の水陸両用車両の点検を行うウェアラブルシステムで、入力を音声で行える。しかし、これらのシステムは個人利用を想定したものであるうえ、位置などのコンテキストと連携したコンテンツの保存、コンテンツ閲覧はできない。

音声ベースのWebページを作成するための記述言語としてVoiceXML[9]がある。また、W3CはVoiceXMLで作成される音声WebページのブラウザVoiceBrowsers[8]に必要とされる機能を規定している。VoiceBrowsersでは電話機によるデータ閲覧も想定し、入力として音声と電話機のプッシュボタンを利用することが考慮されている。出力としては、合成音声によるテキストの読み上げや予め録音しておいた音声データの再生をサポートしている。ユーザはこうした入出力を利用することにより、音声ベースによる情報閲覧を対話的に行なうことができる。しかし、VoiceXMLは空間的な記述などを行うことができないため、既存の音声情報サービス以外への応用が難しい。

SoundWeb[11]は、音声をハイパーリンクの形で構造化し、音声中に埋め込まれたリンクをビーブ音などの副音声で提示する。ユーザはリンクの部分でクリックすることにより、リンク先のコンテンツへアクセスすることができる。SoundWebを利用することで、ユーザはウェアラブルコンピューティング環境において構造化された音声情報を閲覧することが容易となる。また、手軽に音声情報を作成する仕組みも実現されている。しかし、SoundWebは空間的な記述がなく、位置情報と連携することができない。また、ユーザが明示的にコンテンツにアクセスすることを想定しているため、常時利用に適しているとは言いがたい。

SoundShower[13]は、学会のポスター発表などにおいて、その場で話された音声情報をその場に記憶しておき、その説明を聞かなかったユーザや、議論に参加しなかったユーザに対して、説明や議論を提示するシステムである。このシステムでは、ユーザ

のコンテキストを考慮して、音声の聞こえ方（詳細度）を変化させ、ユーザが雰囲気を知ることを可能としている。しかし、SoundShowerは、ある特定の空間のみで情報を共有することを想定したものである。また、音声の空間性については位置以外は考慮されていない。

7 まとめ

本研究では、ウェアラブルコンピューティング環境において、音声のみを用いて記憶共有を図るため、音声情報を空間に記憶するための記述言語を提案し、記述言語をベースとした音声情報システムを実現した。実現したプロトタイプシステムでは、ユーザはマイクなどを利用して音声録音するだけで空間に音声情報を記憶し、ヘッドホンを装着して空間を歩き回ることによって、空間に記憶された音声情報を聞くことができるようになる。ユーザのプロファイル情報に加え、歩行スタイルなどをコンテキストとして利用し、問合せを行うことにより、ユーザに関連性の高いコンテンツを提示するよう工夫している。実現したシステムを利用することで、空間に簡単に記憶を残し、実空間ベースでのコミュニティを作ることが可能となった。

今後の課題としては、記述言語の改良と、ユーザの詳細なコンテキスト取得、複数音声コンテンツを効率よく閲覧するための仕組みを実現することがあげられる。また、音声自体に動作をもたせることも考えられる。さらに、音声の詳細度を切り替えることにより、空間の雰囲気を提示するような仕組みも実現予定である。一方、今回実現したシステムは、ヘッドホンを利用して仮想的に記憶された音声を聴いているだけであるうえ、外部の音をある程度遮断してしまうため、自然環境にある音を大切にしていない。今後は、こうした点を考慮し、マイクなどを利用して直接周囲の音を聞くことができるような仕組みを実現する必要がある。

参考文献

[1] Arons, B.: A review of the cocktail party effect, *Journal of the American Voice I/O Society*, vol. 12, pp. 35-50 (1992).
[2] Kobayashi, M. and Schmandt, C.: Dynamic soundscape: mapping time to space for audio

browsing, in *Proceedings of the CHI '97 on Human factors in computing systems*, pp. 194-201 (1997).
[3] Mann, S.: Definition of wearable computer, Keynote Address for The First International Conference on Wearable Computing ICWC '98(1998).
[4] Rollins, S. and Sundaresan, N.: AVoN calling: AXL for voice-enabled web navigation, in *Proceedings of the Ninth International World Wide Web Conference, Computer Networks*, pp. 533-551 (2000).
[5] Rudnicki, A. I., Reed, S. D. and Thayer, E. H.: SpeechWear: a mobile speech system, in *Proceedings of ICSLP'96*, vol. 1, pp. 538-541 (1996).
[6] Sawhney, N. and Schmandt, C.: Nomadic radio: scaleable and contextual notification for wearable audio messaging, in *Proceeding of the CHI '99 conference on Human factors in computing systems*, pp. 96-103 (1999).
[7] Smailagic, A. and Siewiorek, D. P.: Modalities of interaction with CMU wearable computers, *IEEE Personal Communications*, vol. 3, no. 1, pp. 14-25 (1996).
[8] VoiceBrowsers, <http://www.w3.org/TR/NOTE-voice/>.
[9] VoiceXML Forum, <http://www.voicexml.org/>.
[10] 梅本あずさ, 柴尾忠秀, 水口充, 浦野直樹: 音声提示型インタフェースの実装と評価, インラクティブシステムとソフトウェア VII: 日本ソフトウェア科学会 WISS '99, pp. 83-92 (1999).
[11] 庄司武, 中村聡史, 塚本昌彦, 西尾章治郎: SoundWeb: ウェアラブル環境下でハイパーリンク型音声データのブラウジングを行うシステム, インラクティブシステムとソフトウェア IX: 日本ソフトウェア科学会 WISS '01, pp. 11-16 (2001).
[12] 鳥越 けい子, 今田 匡彦, 若尾 裕: サウンド・エデュケーション, 春秋社 (1992).
[13] Christof E. M., Sumi Y., Mase K.: The Ambient Sound Shower: Experience Sharing with Ubiquitous Sensors, The sixth International Conference on Ubiquitous Computing 2004 (2004).
[14] Rekimoto J., Ayatsuka Y., Hayashi K.: Augment-able Reality: Situated Communication through Physical and Digital Spaces, in *Proceedings of ISWC'98*, p.68 (1998).
[15] 椎尾一郎, 山本吉伸: コミュニケーションツールのための簡易型 AR システム, インラクティブシステムとソフトウェア VIII (日本ソフトウェア科学会 WISS2000) pp. 117-124, 近代科学社 (2000).
[16] 中島 義道: 騒音文化論—なぜ日本の街はこんなうるさいのか, 講談社+α文庫 (2001).
[17] 岩宮 真一郎: 音楽と映像のマルチモーダル・コミュニケーション, 九州大学出版会 (2000).