J-017

人の行動に連動した音を使ったユーザインタフェースの提案と評価
Proposal and evaluation of sound user interface coupled with human behavior

関 洋平† 大谷 拓郎† 岡林 桂樹† 柳沼 義典†
Yohei Seki Takuro Oya Keiju Okabayashi Yoshinori Yaginuma

1. はじめに

ネットワークの高速化，クラウド・コンピューティングの拡大，データマイニング技術等の ICT の発展に伴い，人は様々なサービスを利用することが可能になった．しかし現状では，人がシステムに働きかけなければサービスを利用できないため，人に負担が掛かっている．そこで，能動的にシステム側から人の行動に合わせる，人間中心の考え方に基づくサービス提供へとパラダイムシフトする必要があると考える．

システムが人に情報を提示するとき，人とシステムとの接点（ユーザインタフェース）が重要になる．人にシステムのことを意識させず，人の行動をアシストするユーザインタフェースにより，人に負担を掛けずに多くの価値ある情報を提示可能になる．このような考え方に即した研究として，拡張現実感（AR: Augmented Reality）がある．AR とは，仮想世界上の情報を実世界にマッピングすることによって実世界を拡張するユーザインタフェース技術である．これにより，情報とモノの関係性把握が容易となり，効率良く情報を享受できると期待されており，既に観光の現場などで使われ始めている[1]．しかし，AR の研究の多くは視覚を利用しており，聴覚を利用したものは少ない[2]．また，情報提示手段として聴覚を利用することで，目や手を他のタスクに割り当てながら価値ある情報を提示できる．

そこで本研究では，仮想世界上の情報を実世界にマッピングして提示する手段として，発展の余地が大きいと期待する聴覚の特徴を活かした音声インタフェースを提案する．提案した音声インタフェースは，人の頭部姿勢に連動してリアルタイムに音像定位を行う特徴を持ち，人の注視行動とその注視している時間に応じて複数音源から興味ある音源を徐々に絞り込むことが可能である．本インタフェースを試作し，音像定位の性能と音源の絞り込みに対する有効性を実験により評価した．

2. 聴覚の特徴を活かした情報提示

聴覚には，あらゆる方向の音を知覚可能な全方位性があり，自分の周囲のどの位置に音源が存在するか認知することが可能である．そのため，聴覚の全方位性を利用した音声インタフェースにより，マッピングされた音声情報の位置を捉えやすくなると考えた．

人は日常生活の様々な場面で聴覚の全方位性を活かしている．例えば，バーゲンセールなどで，各店舗が周囲に向けて宣伝をしているとする．そのとき，ユーザの動きに連動して，周囲にある各店舗の方向からその店舗の宣伝をイヤホンに再生する（図 1）．これにより，自分の周囲のどこにどんなバーゲンセールをしている店舗があるのか把握が容易となり，効率良く興味がある店舗を探すことができる．この実現には，音の方向を仮想的に作り出し，人に音源の位置を認識させる音像定位技術が必要となる[3]．音像定位技術を利用したシステムには，視覚障碍者の移動支援システムなどがある[4]．さらに，頭部姿勢に追従して音像定位を行うことで，現場を忠実に再現するヘッドホンシステムがある[5]．しかし，いずれも特定の場面に用途が限定されている．日常生活の様々な場面で聴覚の全方位性の特徴を活かすには，実世界を自由に動き回るユーザが，情報を積極的に探索することも考慮して設計する必要がある．

一方，聴覚の特徴として，同時に提示された複数音源から興味ある音源だけを選別して聴取可能なことは，カクテルパーティー効果として知られているが，複数音源から興味ある音源だけを集中して聴くため，人に掛かるストレスが大きくなる．このような問題を解決するために梅津らの提案では，頭部姿勢のみに注目し，ユーザの頭部の向きに応じた音量制御を行っている[2]．しかし，ユーザは探索行動するとき様々な方向を見るため，必ずしもユーザの頭部の向きが常に興味ある音源を示しているとはいえない．そのため，対象物にどのぐらい興味を持っているか判断するには，人の行動だけでなく，注視時間といった時間的要素の考慮も必要と考える．また浜中らのシステムは，意図的にハンドジェスチャを行う必要があり，またそれにより手を自由に動かせなくなるため，ユーザの負担になっている[6]．

そこで本研究では，人の注視行動に注目し，注視時間に伴ってユーザが興味ある音源を自然に絞り込むユーザインタフェースを考案した．これにより，ユーザが今何に注目しているかを，時間的要素を考慮した行動から読み取り，それに応じてダイナミックに方向付けした音声情報の提示が可能になると考える．

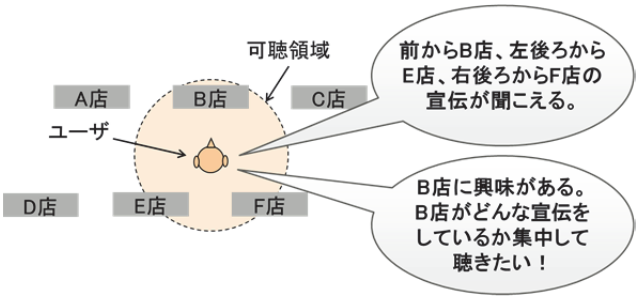

図 1: 聴覚の全方位性を活かした場面の一例

3. 注視行動に連動したユーザインタフェース

例えば，図 1 のような場面で，人が何か興味あるモノを探すとき，人は移動・探索・注視を繰り返すと考え，行動モデルを立案した（図 2）．まずは移動し，得たい情報を動き回りながら探索する．そして，ある情報を見つけたときそれが得たい情報なら，そこで注視し，不必要な情報な

† （株）富士通研究所 FUJITSU LABORATORIES LTD.

ら再度，移動・探索を行う．我々はこの注視した対象をユーザの興味ある対象と捉え，注視行動を利用して音声情報を提示する注視ユーザインタフェース（以降，注視 UI）を考案した．注視 UI とは，複数音源が存在する環境下で，人が音源の選択を意識せずに，注視行動と連動して聴きたい音源を徐々に絞り込む，人の行動と時間的要素を組み合わせたユーザインタフェースである．注視行動は，ユーザの頭部姿勢から認識可能である．これによりユーザは，システムの存在を考えず，自然な動作で音源の取捨選択が可能になる．

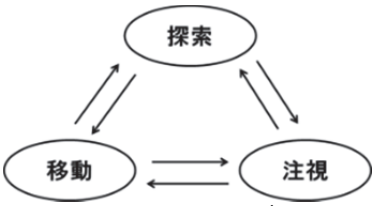

図 2: 人の行動モデル

注視 UI の処理の流れを以下に示す（図 3，図 4）．可聴領域とは，音源が聞こえる範囲のことを指す．

1. ユーザの頭部姿勢角度からユーザの見ている方向を判定する
2. ユーザが同じ方向を Ta 秒以上見続けたとき，注視行動と判定する
3. 注視行動と判定してから，ユーザが同じ方向を見続けている間，可聴領域を収束角度 θt（ユーザの顔正面を 0 度とする）まで徐々に小さくする
4. 可聴領域を小さくするとき，収束角度以外にある音源はフェードアウトさせる（図 5）
5. 可聴領域が小さくなっている間，頭部姿勢角度がある閾値以上変化したとき，可聴領域を元に戻す

なお本稿では，注視行動と判定するまでの時間 Ta を 2.7 秒，収束角度 θt を±30 度に設定した．

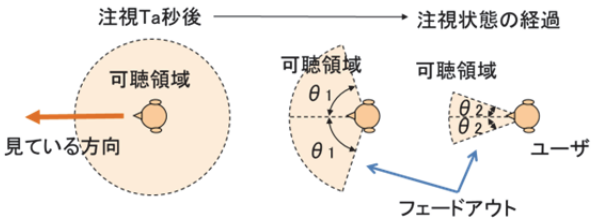

図 3: 注視 UI の仕組み

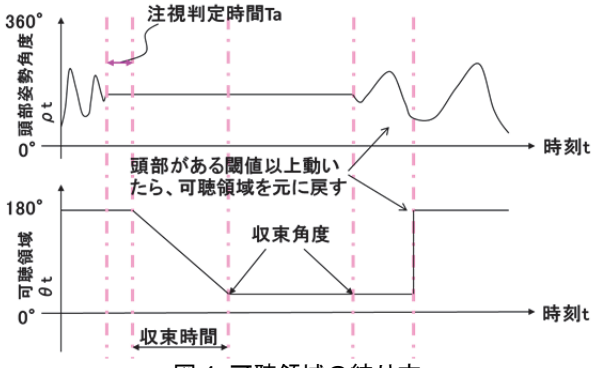

図 4: 可聴領域の絞り方

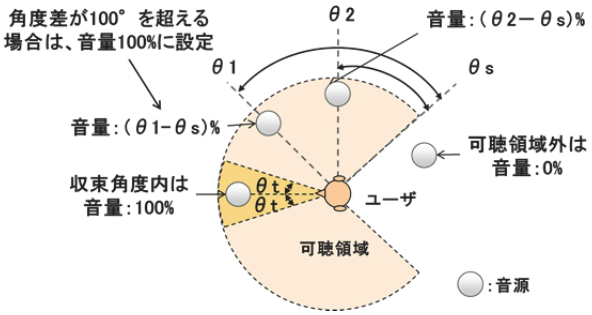

図 5: フェードアウトの仕組み

4. システム概要

本システムは，ノート PC（OS: Windows 7），イヤホン（オーディオテクニカ: ATH-CK70PRO），角度センサ（Trivisio Prototyping GmbH: Trivisio Colibri）で構成されている（図 6）．角度センサは，加速度，ジャイロ，磁気センサを用いている．頭部の角度センサから取得されるユーザの頭部姿勢角度に基づき，音声データに音源から両耳までの音の伝達特性（HRTF: Head-Related Transfer Function）を PC 上で畳み込むことで音像を定位させて，イヤホンに出力する．HRTF は，MIT メディアラボが一般に公開しているものを使用した[7].

また，顔前方の音像の定位感向上を図るために，後方の音源より前方の音源の音量を相対的に大きくする強調処理を施した．

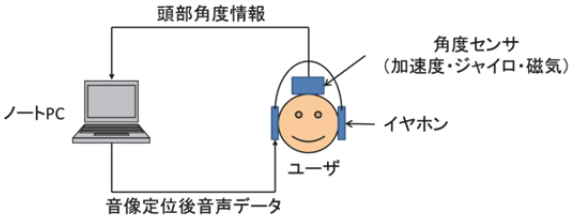

図 6: システム構成

5. 音像定位性能実験

HRTF，頭部姿勢のセンシング精度，センシングを含む音像定位のレスポンスにより，音像の定位感が異なるため，本システムにおける音像の定位感を観察する必要があると考えた．

また，人が聞こえてくる音の方向を頼りにして，注視行動を行うためには，高い精度の音像定位が必要である．本システムのように頭部姿勢と連動させて音像を定位する場合は，より高い精度での音像定位を実現できると推測する．そこで，注視 UI の有効性を評価するにあたり，まず基礎実験として音像定位の精度を絶対精度と複数音源の分離という観点で評価した．

5.1 頭部姿勢に連動した音像の定位感実験

5.1.1 実験目的

頭部姿勢と連動させた音像定位が，定位感向上をもたらすことが知られており[3]，本システムにおいても，同様の効果が見られるか検証した．

5.1.2 実験方法・条件

音源の提示角度条件ごとに定位させた音源を実験協力者に聞かせ，音源の方向を 1 度目盛間隔の円周が描かれたシートにプロットさせる．実験条件と音源の内容を表 1 に示す．頭部運動条件は，頭部姿勢に連動して音像定位を行う．各実験協力者に頭部静止条件と頭部運動条件の両条件を行わせる．

表 1: 実験条件と音源の内容（顔正面を 0 度とする）

頭部姿勢条件: 2 条件	頭部静止条件 ・頭部を静止 ・角度センサを取り付けない	頭部運動条件 ・±45 度以内で頭部を左右に動作可能
音源の提示角度条件: 24 条件	右回りに 360 度を 15 度間隔区切りにした角度（図 7） 1 条件 1 回，ランダムに提示	
音源の内容	ニュース原稿を読み上げた男性の音声，10 秒	

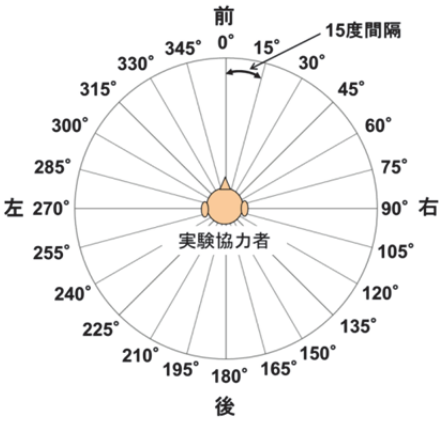

図 7: 音源の提示角度

5.1.3 実験結果

実験協力者は 8 人（20 代～50 代の各世代 2 人ずつ）で，本システムの仕組みを知らず，初めて使用する人である．350 度と回答した場合は，-10 度と回答したとみなすように，適宜 360 度の補正をした．全実験協力者の回答角度の平均結果を図 8（頭部静止条件），図 9（頭部運動条件）に示す．ひげ線は標準偏差を表す．

図 8，図 9 より，頭部静止条件における横方向（右方向 30 度～150 度，左方向 210 度～330 度）の音源提示では，右方向は 90 度付近，左方向は 270 度付近に平均回答角度が分布している．それに対して頭部運動条件では，真値付近に平均回答角度が分布している．ただし，標準偏差は頭部運動条件の方が大きい．

また，頭部静止・頭部運動の両条件において前方向（0 度）で標準偏差が大きい．これは前方の音源を後方と誤認識した回答が多いためである．さらに，ある真値とその真値に対する全実験協力者の平均回答角度から二乗平均誤差を算出したところ，頭部静止条件は 37.02 度，頭部運動条件は 27.61 度であり，頭部運動条件の方が真値からの誤差が小さいことがわかった．

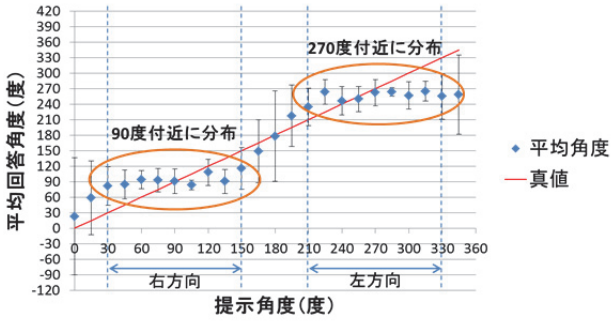

図 8: 頭部静止条件の回答角度の平均

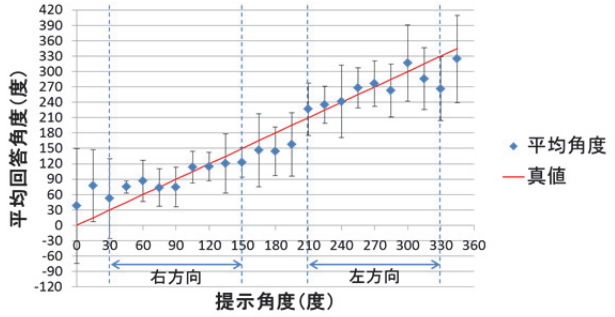

図 9: 頭部運動条件の回答角度の平均

5.2 2 音源の分解精度実験

5.2.1 実験目的

異なる方向にある 2 つの音源を同時に再生させたときに相対的に 2 音源を分解可能か検証した．ここで，目標分解角度を設定するにあたり，本システムの利用場面の一例として展示会場を採り上げ，展示会場の場面を図 10 のように想定し，目標分解角度を 25 度（≒24.7 度）と算出した．

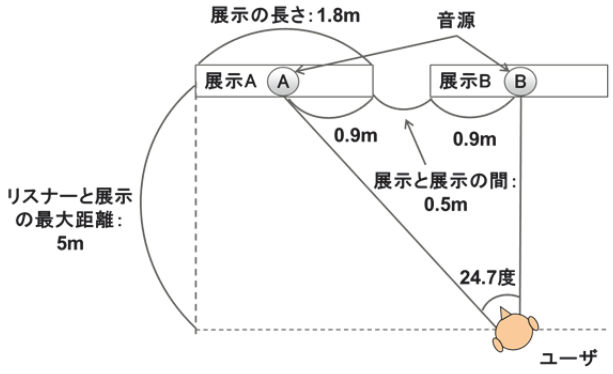

図 10: 展示物の想定配置

5.2.2 実験方法・条件

5.1 節における頭部姿勢に連動した音像の定位感実験では，前方の音源を後方と誤認識する回答が多くあった．しかし，音源位置を前後で誤認識すると，2 音源の相対的な位置関係の確認ができない．そこで，音源位置の前後の誤認識を軽減させるため，事前に実験協力者に「音源は顔正面に近づくほど音量が大きく再生され，真後ろに近づくほど音量が小さく再生される」といった本システムの説明をし，4 分間システムを使用させ慣れさせた．4 分間という

学習時間は，実験協力者 2 人による熟練度評価実験により決定した（表 2，図 11）．

実験では，条件ごとに定位させた 2 音源を同時に実験協力者に聞かせ，2 音源の方向がわかった時点で再生停止ボタンを押させる．そして，円周が描かれたシートにその 2 音源の方向をプロットさせる．頭部は自由に動かして良いとした．実験条件と音源の内容を表 3 に示す．

表 2: 熟練度評価実験の実験条件と音源の内容

回答方法	音源の方向を 1 度目盛間隔の円周が描かれたシートにプロット
音源の提示角度条件: 8 条件	右回りに 360 度を 45 度間隔区切りにした角度 1 条件 1 回，ランダムに提示
システムの仕組みの説明教示・練習時間	教示無：0 分 教示有：0 分，1 分，2 分，3 分，4 分，5 分
音源の内容	ニュース原稿を読み上げた男性の音声 最長 1 分

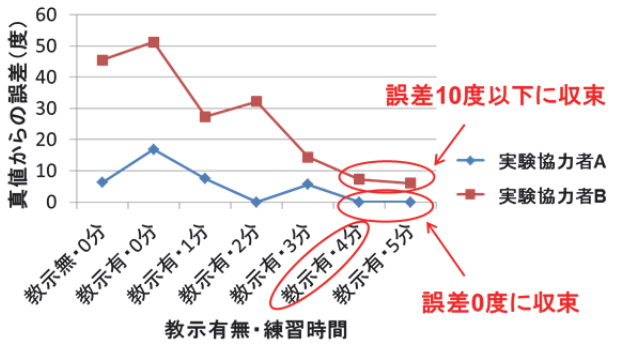

図 11: 熟練度評価実験結果（真値と回答角度の誤差の平均）

表 3: 2 音源の分解精度実験の実験条件と音源の内容

2 音源の間隔角度条件: 4 条件	20 度，22.5 度，25 度，35 度（図 12） 1 条件 8 回提示，ランダムに提示
2 音源の内容	異なるニュースの原稿を読み上げた男性の音声と女性の音声，最長 1 分

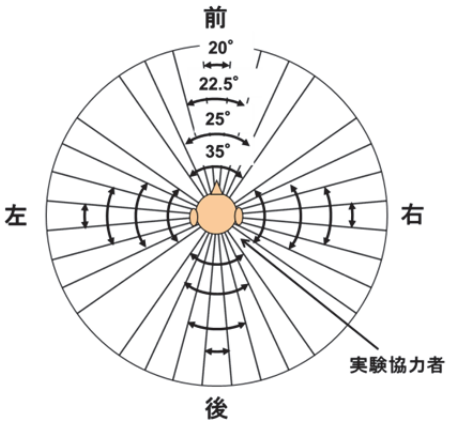

図 12: 2 音源の配置位置

5.2.3 実験結果

実験協力者は 8 人（20 代: 3 人，30 代: 2 人，40 代: 3 人）である．2 音源の相対的位置関係（男性の声が女性に対して左右どちらにあったか）の正答率について，全実験協力者の平均を図 13 に示す．

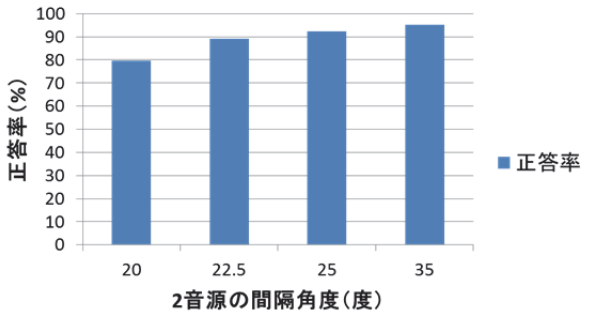

図 13: 2 音源の相対的位置関係の正答率

正答率に対して，片側二項検定を行った結果，全ての間隔角度条件で正答率が有意水準 1%で有意に高かった．この結果より，本システムは目標分解角度 25 度で 2 音源を分解可能と確認できた．

5.3 音像定位性能実験の考察

頭部運動条件が頭部静止条件より，音像が真値付近に定位することが確認できた．頭部静止条件では，横方向（右方向 30 度～150 度，左方向 210 度～330 度）の音源提示において，例えば 30 度に定位させた音源を 90 度と認識するように，斜め方向の音源を真横（90 度，270 度）と認識する傾向があり，バラつきは少ないが，定位感は低かった．それに対して，頭部運動条件では，バラつきはあるものの斜め方向の音源がその真値付近にあると認識する傾向があった．このことから，頭部姿勢連動の音像定位により，横方向の音源の位置が認識しやすくなったといえる．

一方，前方の音源を後方と誤認識する傾向は，先行研究でも示されており[8]，本システムにおいても同じ現象が観察された．前方の音源の音量を大きくする強調処理を施したが，その仕組みの説明なしでシステムを実験協力者に利用させた場合は，ただ音量が大きくなると感じるだけで，音源が前方にあるとは認識しなかった．ただし，システムの仕組みを説明し，4 分間システムを使用させることで，音源位置の前後の誤認識がほとんどなくなった（図 11）．これは，強調処理の仕組みの理解と体験による学習効果で，音源の前後の判断を正しく判断できるようになったものと考えられる．

6. 注視 UI の有効性の評価実験

6.1 実験目的

複数音源がある環境下で，注視 UI がユーザに興味ある音声を自然に絞って提供できるか検証した．ここでは，2 音源の分解精度実験と同様に，一例として展示会場の利用場面を想定し実験を行った．

6.2 実験方法・条件

3 種類の展示パネルとその説明を読み上げた音源（男性の音声×2，女性の音声×1）を用意する．実験協力者を中心として，周囲に 3 つの展示パネルを内側向きに置く．

実験協力者には，その中心付近で自由に体を動かして良いと指示し，音声を聞かせる．また，「新しい展示案内システムを 2 つ開発したので（注視 UI あり・なし），両システムを聴き比べてもらう」，「展示パネルの方向から展示の説明音声が聞こえるシステムである」と教示する．

実験条件は，注視 UI あり・なしの 2 条件である．1 回目に注視 UI なし，2 回目に注視 UI ありのシステムをそれぞれ 3 分間ずつ使用させ，最後にアンケートを行う．実験環境を図 14，実験の様子を図 15 に示す．

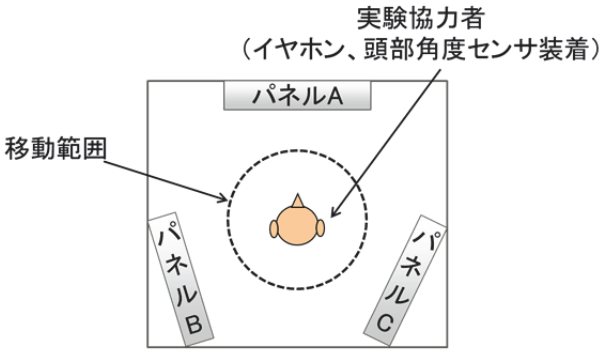

図 14: 実験環境

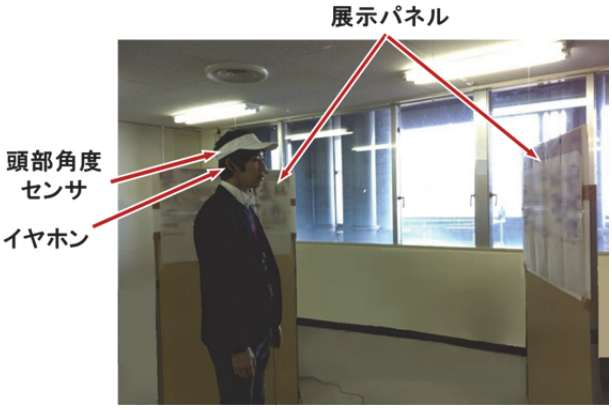

図 15: 実験の様子

6.3 実験結果

実験協力者は 8 人（20 代: 2 人，30 代: 4 人，40 代: 1 人，50 代: 1 人）に対し，主観アンケートを行った．

まず，1 回目（注視 UI なし）と 2 回目（注視 UI あり）で，どちらのシステムが展示の説明を聴きやすかったか質問したところ，全実験協力者が注視 UI ありを選択した．理由の自由記述では「1 回目は，様々な音声が聴こえてうるさく，聴きたい音声が聴き取れない．」，「2 回目は，注視した展示の音声のみが聴こえ，周囲の音声がカットされるため，聴きたい展示に集中できたから．」といった回答が全実験協力者から寄せられた．このことから，注視 UI はユーザが興味ある音声を絞って提供できることが示唆された．

注視 UI ありのシステムに対して，全実験協力者が音を絞り込む機能に気づいたので，次に「質問 1: 音の絞り込みは自然だったか？（5 を最も自然，1 を不自然とする 5 段階評価）」，「質問 2: 音の絞り込みのタイミングは適切だったか？（5 を最も適切，1 を不適切とする 5 段階評価）」を尋ねた．それぞれの各回答率の平均を図 16（音の絞り込みの自然さ），図 17（音を絞り込み始めるタイミングの適切さ）に示す．

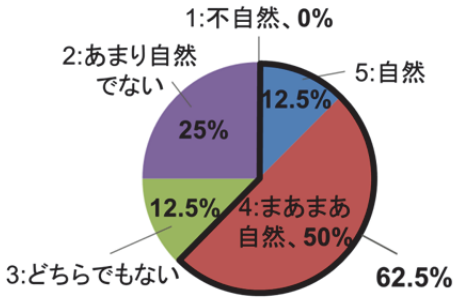

図 16: 音の絞り込みの自然さ

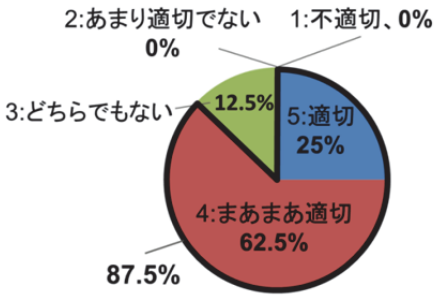

図 17: 音を絞り込み始めるタイミングの適切さ

図 16，図 17 より，実験協力者の 62.5%が音の絞り込み方が自然と感じ，87.5%が音を絞り込み始めるタイミングが適切と感じたことがわかった．これにより注視 UI は，興味ある音声を自然な絞り方，適切なタイミングで提示できたと考えられる．

さらに参考として，美術館などに設置してある選択式の展示案内システムと本システムの良さを比較した．ここで選択式のシステムとは，手に持った端末に展示番号を入力して展示物を選択したり，無線タグにタッチして展示物を選択したりして展示物の説明音声を聴くシステムのことを指す．「選択式の展示案内システムと比較して今回の展示案内システムはどうだったか？（5:今回の方が良い，4:どちらかと言えば今回の方が良い，3:変わらない，2:どちらかと言えば選択式の方が良い，1:選択式の方が良い）」をそれぞれ注視 UI なしと注視 UI ありで尋ねた．選択式のシステムを利用したことがない人は，想像して回答させた．各回答率を図 18（注視 UI なしとの比較），図 19（注視 UI ありとの比較）に示す．

図 18，図 19 より，選択式のシステムと比べ，注視 UI なしは 12.5%の実験協力者しか良いと評価しなかったのに対し，注視 UI ありでは 75%の実験協力者が良いと評価した．

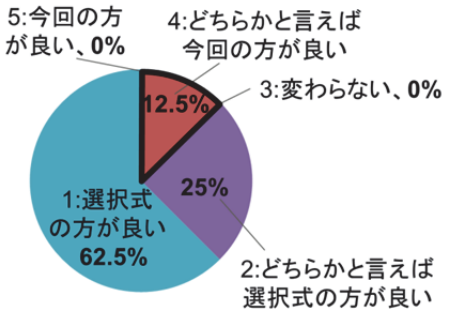

図 18: 注視 UI なしと選択式の比較

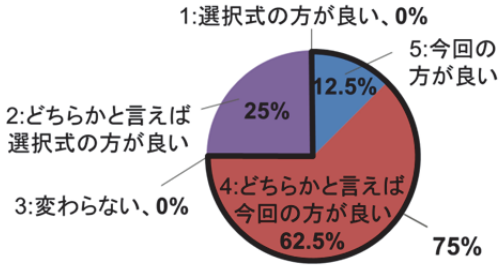

図 19: 注視 UI ありと選択式との比較

6.4 注視 UI の有効性の評価実験の考察

注視 UI あり・なしの比較では，全実験協力者が注視 UI ありは展示の説明を聴きやすいと回答しており，ユーザの注視行動に応じて，興味がある音声だけを絞って提示する注視 UI の有効性が示された．しかし，注視 UI の音の絞り込みの自然さにおいて，あまり自然でないと感じる人が一部いた．理由を尋ねたところ，音を絞り込む速度が遅いといった意見があった．本システムでは，音を絞り込む速度やタイミングの設定値を数人の主観評価で決定したため，現段階では最適値になっていない．そのため今後は，実験によって設定値を調整する必要がある．

注視 UI なしより選択式のシステムを良いと感じた人が多かった理由は，聞こえてくる興味ない音声がノイズとしてストレスになったからと考えられる．それに対し，選択式より注視 UI のシステムがより多く選ばれたのは，最初に自分の周囲にどんな音声情報があるのか知ることができ，その中から興味を持った音声を自然に絞ることができたからと考えられる．

しかし，注視 UI より選択式のシステムが良いと感じた人が一部いた．注視 UI ありのシステムでも最初は全ての音声が聞こえ，それがストレスに感じるといった意見が寄せられた．選択式のシステムは，ユーザの興味の対象が最初から決まっているときには使い易いが，何の情報もない状態から興味ある対象を探すときは，1 つずつ選択して確認する必要があり不便である．本実験では，用意した展示パネルが 3 つだけであり，また実験協力者の周辺近くに配置してあったため，説明が聴きたい対象をあらかじめ決めやすい状態であった．そのため，選択式システムで 1 つずつ対象を選択したいと感じた人がいたと考えられる．一方，注視 UI ありのシステムは，同時に複数の音声情報をユーザに提示できるという利点を残しつつ，その中から興味ある音声情報のみを自然に絞ることが可能である．以上のことから，ユーザが自ら探索し，興味ある情報を得ようとする場面などにおいて，注視 UI は有効であると考えられる．

7. まとめ

本研究では，ICT を利用した人間中心の考え方に基づくサービス提供を目指し，能動的にシステム側から人にサービスを提供する場面で重要となる，人とシステムの接点として，新たな音声インタフェースを提案した．提案した音声インタフェースの音像定位の性能評価と注視 UI の有効性の評価を展示会場の場面を一例として想定して行った．その結果，以下のことが明らかになった．

- 頭部姿勢連動は音像の定位感向上に効果がある
- 25 度間隔の 2 音源を聞き分けられる
- 注視 UI は，興味ある音源を自然に絞って提示可能

この結果より，人の行動と時間に応じてダイナミックに音声情報を提示することは有効であると考えられる．

本研究では，注視行動のみに注目してユーザインタフェースを設計したが，今後は探索や移動といった他の行動にも応じて，効率良くユーザに情報を提示する仕組みを考える．また，前方の音像定位精度の向上も試みる．音像にわずかな運動を加えることによって音像定位精度が向上することが示されており[9]，この仕組みを本システムに組み込むことで定位精度の向上が期待できる．さらに本システムは，MIT メディアラボが公開している HRTF を利用したが，個々人に合った HRTF を個人ごとに適応することでも，定位精度の向上が可能と考える．

本実験で得られた結果を基に，人が使いやすいシステムに改良し，本システムが人とシステムの接点を増加させ，ICT が人の生活を今まで以上にサポートできるようになることを目指す．

参考文献

[1] 藤本義治, 小野哲雄, “拡張現実感を用いた新たな観光パンフレットの提案”, 情報処理学会創立 50 周年記念（第 72 回）全国大会講演論文集, Vol.72, No.4, pp.4.437-4.438, (2010)
[2] 梅津直貴，井ノ上寛人，堀内恒，佐藤美恵，小黒久史，春日正男，“空間把握性に注目した音響案内システムの開発に関する研究”，映像情報メディア学会技術報告，Vol.35, No.39, pp.41-44, (2011)
[3] 平原達也, “バイノーラル信号による音像定位技術 -動的バイノーラル信号の効用-”, 騒音制御, Vol.33, No.3, pp.204-211, (2009)
[4] 山本雅大, “音像定位技術を用いた視覚障碍者の移動支援”, 第 2 回トロン/ユビキタス技術研究会, (2010)
[5] ソニー(株), MDR-DS7500, “http://www.sony.jp/headphone/products/MDR-DS7500/”, (2012/4/10 現在)
[6] 浜中雅俊, 李昇姫, “サウンドスコープヘッドフォン”, 日本バーチャルリアリティ学会論文誌, Vol.12, No.3, pp.295-304, (2007)
[7] MIT Media Lab, HRTF Measurements of a KEMAR Dummy-Head Microphone, “http://sound.media.mit.edu/resources/KEMAR.html”, (2012/4/10 現在)
[8] 村井厚介, 小川和宏, 降旗建治, 柳沢武三郎, “耳栓型イヤホンによる頭外音像定位の可能性”, 電子情報通信学会技術研究報告. EA, 応用音響, Vol.102, No.398, pp.69-74, (2002)
[9] 森田貴大, 李周浩, “知能化空間における立体音響を用いた人物誘導(第 2 報) -定位認識向上のための音像運動パターンの提案-”, 第 12 回計測自動制御学会システムインテグレーション部門講演会, 3D3-2, (2011)