

Title	視覚情報を補完する静穏並列型音情報システム
Author(s)	宮地, 泰造; 田窪, 勇人; 渡辺, 正介
Citation	第六回知識創造支援システムシンポジウム報告書: 196-204
Issue Date	2009-03-30
Type	Conference Paper
Text version	publisher
URL	http://hdl.handle.net/10119/7987
Rights	本著作物の著作権は著者に帰属します。
Description	第六回知識創造支援システムシンポジウム, 主催: 日本創造学会, 北陸先端科学技術大学院大学, 共催: 石川県産業創出支援機構文部科学省知的クラスター創成事業金沢地域「アウェアホームのためのアウェア技術の開発研究」, 開催: 平成21年2月26日~28日, 報告書発行: 平成21年3月30日

視覚情報を補完する静穏並列型音情報システム

A Quiet parallel Voice Navigation System for Complementing Visual Information

宮地 泰造 東海大学情報理工学部情報メディア学科
 Taizo Miyachi School of Information Science and Technology, Tokai University
 miyachi@keyaki.cc.u-tokai.ac.jp

田窪 勇人 (同上)
 Hayato Takubo 5adm3205@keyaki.cc.u-tokai.ac.jp

渡辺 正介 (同上)
 Shosuke Watanabe 4adm3110@keyaki.cc.u-tokai.ac.jp

Keywords: voice navigation, digital signage, parametric speaker, parallel listening, visiting business

Summary

Navigation and assist for human actions by digital signage systems supplies visitors with well arranged knowledge that can not easily focusing by internet search engines. Voice navigation that complements deep knowledge in visual information by digital signage allows visitors not only to choose but to acquire the knowledge in a limited time and a limited space without sound pollution in order to expand international tourism exchange and domestic demands. In this paper we propose a methodology by parametric speakers for easy parallel listening of explanation with long sentences based on potential power of human brain.

1. はじめに

世界の観光産業は、日本の自動車産業よりも規模が大きく、観光産業は観光立国日本を目指す我国にとって重要である。観光産業は、裾野が広く、観光施設以外に、ショッピング、飲食店、交通機関、宿泊、コンサート、スポーツ、動植物園などがある。

ユビキタスコンピューティング時代に突入した今、検索エンジンを介して大量情報を収集可能になっているが、目的達成に必要な情報収集には多くの時間が必要であり、必要なすべての情報を得られる訳ではない。目的地や環境も時々刻々と変化している。

多くの場合人は、有名な情報や手掛かりの事前情報を得て、目的地に行く。目的地で欲しい情報や多種類の有用な情報を入手する。目的地で初めて分かる情報はたいへん重要であり、事前情報の価値評価と不足情報の補完を一挙に行うために有効である。訪問地で実際の意味が理解でき、移動中の短時間に行動支援の知識を提供できることから、デジタルサイネージ[Rose 06]が重要になってきている。有名な観光地や人気のモールでは、たとえば、人気のエキュート品川[ecute 08]は、駅中という最高の立地条件

で、誰もが納得する最高の商品とサービスに加え、定期的な魅力店舗の入れ替えにより、リピーターを獲得している。

観光立国日本や内需促進に重要なことは、必要な時、必要な場所/空間で、状況・コンテキストに合った知識・情報を、センサー情報を含めて絞り込み、提供することである。デジタルサイネージの視覚情報と、その中に隠れている意味や視覚情報に跨る知識を、移動中の限られた時空間で分かり易く提示することにより、行動が豊かになる能動的知識循環社会を構築できる。人間は、周囲 360 度の空間から、音により同時に複数の情報を得ることができ、意識しなくても自然に、必要な情報を選別している。

本稿では、視覚情報の深い意味や複数の関連情報を、同時に聞き取り・選択できる人間に対して、騒音を出さない超音波により複数情報を並行して提供する音情報提供システムとその提供手法について、人間の脳の限界処理という観点も含めて、得られた知見を報告する。

2. 目的地での知識の編み込み

世界中にユニバーサルな知識・情報の共有が進み、多様化した個人の嗜好に合う、感動やリスク回避などの経験交換により、高い満足度を得るための知識

共有環境がインターネット上に形成されつつある。

しかし得られた知識・情報のとおりにはすべてが進む訳ではないため、個々人は目的地での新たな発見や興味対象の変化に対し、自分の嗜好に合わせて知識を編み込んでいくことが重要になる。

2.1 事前情報、現地情報と知識の編み込み

一般的に、賢明な人の経験や知識を、事前知識に編み込むことにより、効率的に高い満足を得られる。人間は毎日動いており、静止していない時間がむしろ多い人もいる。そこで、移動中の限られた時間・空間で、多様な嗜好に対して有用な知識情報や気づきを、必用な所で、必要な時に、必要なだけ、センサー情報も含めて、提供することが重要になる。

たとえば、目的地に行き、インフォメーション：「i」で情報を得ることにより、事前情報の有効性が明らかになるとともに、散在していた事前知識が繋がってくる。たとえば、城とスポーツイベントの情報を事前に持っている場合、現地でスポーツイベント関連展示会や城近辺の遊覧船や散歩コースがあることに気づけば、これらを繋いで一日の観光プランを作り、嗜好を反映して一日を優雅に楽しめることになる。事前情報を縦糸にし、現地情報を横糸として、自分の嗜好に合うプランを編み込むことができる。この背景には、城のある丘と運河の自然公園の

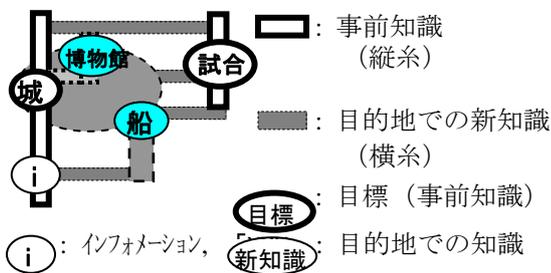


図1. 事前知識と 目的地の知識との編み込み

隣接関係に気づき、さらに、現地の文化や有名な食べ物の発見、現地での人とのコミュニケーションと一緒に時間を過ごす楽しみの発見があり、これらを自分流に編み込んでいくことにより、高い満足感が得られることになる。

2.2 事前/現地情報と脳知、共鳴知、実践知

事前知識は、頭で認識した「脳知」であり、現地では自然環境やイベント・文化に基づく「共鳴知」、「シーン知」、「環境知」、「実践知」が重要になる。人間は、知識の編み込みを、共鳴知、シーン知、環境知、実践知を基に行うと考えられる。

(1)共鳴知とは、対象物やその環境に対する、人の心

や体の共鳴に基づく知識、および、潜在的な志向や嗜好への気づきである。人が対象物やその環境に対する知識に対して、何らかの共鳴を感じて、価値と志向性を認めたとき、これを「共鳴知」と呼ぶ。対象物は何でもよく、環境の例として、現地の自然環境・施設環境・交通環境・文化環境・人環境などである。

観光やショッピングの馴染みの少ない場面や有効な解決策が必要な場合が多く、自分の志向の向きを仮に決めて、良い手掛かりを早く獲得したいという気持ちが働く。そこで人間には、ある程度、納得がいく知識・情報である共鳴知の入手が重要になる。その結果、共鳴知は、印象的で鮮明であり、自然と納得する共鳴点や論理性があることになる。

人間は、共鳴した方向に、苦勞すること無く自然に志向・論理全体の向きを変えるため、その向きに従って「想像活動（希望的イメージ作り）」を行う。共鳴知は、形式知と確信のある暗黙知のほか、制約が少ない点から評価される暗黙知も含んでいる。共鳴知は、目標や志向の方向を決めるため、共鳴知に従って、「知識の収集・分析・獲得・活用」が進む。活用では、行動立案などの「創造活動」を行う。

(2)共鳴知は、「シーン知」を伴う。城には、美しい外形、偉大なる王の間、豪華な宴の間、一流品の贈り物が一同に並ぶというシーン知がある。シーン知の表出により目的対象が想像可能になる。共鳴知はシーン知を気づかせ、シーン知は関連するシーンでの新たな関連シーン知の発見に繋がる。鎧・兜・武器に見られる騎士団の勇猛さ、芸術品に見られる西洋と東洋の文明交流、優雅な時間を過ごせる一流レストランやカフェなどが関連シーン知を構成する。関連シーン知は、新たな共鳴知の発見に繋がる。この過程では、形式知への気づきだけでなく、よく体験する暗黙知への気づきも働く。多くの気づきの提供により、価値のある共鳴知が増えて、共鳴知の集合はより高い価値を持つ。



図2. 脳知、共鳴知とシーン知のサイクル

人は、環境が変化すると、満足する対象、行動内容および満足度も異なる。共鳴知とシーン知は、「環境知」を伴う。シーン知は環境知を気づかせ、環境知は、さらに拡大した環境の拡大環境知の発見に繋がる。拡大環境知は、環境に対する環境共鳴知を表出させる。環境知の特徴は、「制約」記述が存在する

点である。人間の心と体が共鳴点を無意識に見つけられる環境[Gibson 66]を「共鳴環境」と呼ぶ。共鳴環境は多様な要素を含んでおり、人間の共鳴知は、個人ごとの「潜在的共鳴知」と新たに気づいた「発見的共鳴知」を発現して知識の編み込みを行う。

環境知の例は、美しい庭園のある宮殿、貿易品を運ぶ川、一流のブランド店が並ぶフロアがある。

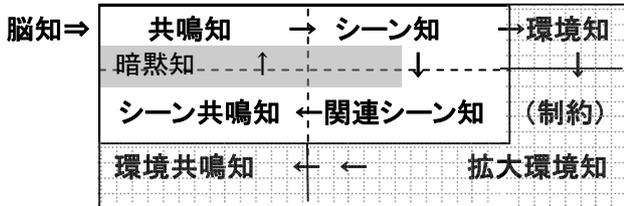


図3. 共鳴知、シーン知と環境知のサイクル

拡大環境知には、難攻不落の眺めの良い城、東洋と西洋の中心中継地、優雅な川下りの景勝地、駅を出なくても乗換えや出勤途中に立寄れる場所がある。(3)想像活動に続いて、創造活動や行動を実践する場合に、実際に行ってみて明らかになる「実践知」がある。共鳴知により選択された志向や目的を実践する実践知、最初の実践により新たに発見・発展した目標を行う発展型実践知、発展型の実践により、新たに実践共鳴知が見つかる。実践知の特徴は、解決すべき「問題」とその「解決案」の記述である。

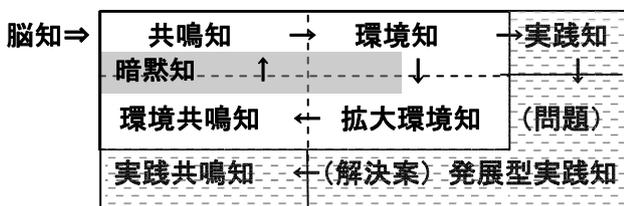


図4. 脳知、共鳴知と実践知のサイクル

3. デジタルサイネージの情報提供

現地での知識の編み込みの主な条件は、つぎの6つである。

- (a) 必要な場所で、必要な時に、共鳴知が得られる
- (c) 移動中の限られた時空間で深い知識が獲得可能
- (d) 自分の嗜好に合わせて気づくことができる
- (e) 難しい操作をしなくてすむ
- (f) 容易に知識を想像、理解できる
- (g) 容易に知識を組み合わせて、立案できる

利用者状況に適応して、適切な共鳴知、シーン知、環境知、実践知の関連つけた提示が有用になる。

3.1 ポスターとデジタルサイネージ

人間は、高い壁にあるポスターやデジタルサイネージを見ることにより、新たなライフスタイルなどに気づくことができる。デジタルサイネージは、インターネットに繋いで、リアルタイムで連携させることも可能である。しかし、一般的なデジタルサイネージ(例えば”UBWALL” [Fujitsu General 08])は、通路や広場の横に設置されるため、横を通過する短時間での情報提供となってしまうため、通行人には情報入手が容易ではない。

3.2 サインによるメッセージ

一般にポスター中にメッセージは、絵・写真やキャッチフレーズを表す飾り文字の形で表示する。しかし、通行人の足を止めるほどの深い感銘を与えることは容易ではない。視覚的メディアにより心理的に好い印象を与えるために、認識、注目、記憶、感情、嗜好、連想の視点からの手法[Franconeri 05]がある。サインにより、深い感銘を与えるだけでなく、サイン・システムとのインタラクションを開始できるようにすべきである。サイン・システムは、利用者に分かり易いメッセージにより、行っている行動に対する有用な知識・情報を環境からのアフォーダンスとして提供し、行動の選択に確信を得られるようにする。

案内の観点から、サインは主に4つに分類できる。

- (i) 印象, (ii) キーワード/キャッチフレーズ, (iii) 長文による説明, (iv) 臨場感のあるストーリー である。
- (i) 印象: サイン・システムは、人々の注意(attention)を獲得できなければならない。そのために、魅力的な視覚表現として、印象的な色使いやカラーコーディネーション、感情あふれる写真、飾り文字などを使う。
- (ii) キーワード/キャッチフレーズ: 人々は、ポスター内のキーワード/キャッチフレーズから、新しい発見、楽しい体験、重要な行動に、気づくことができる。壁の大きいポスターや巨大デジタルサイネージは見易いが、通路やフロアに設置されている場合は、周囲の人に視線を遮られて情報が入手できなくなる
- (iii) 長文による説明: 現地で不足情報を入手するとき、整理済みの分かり易い長文による詳細情報が有用であり、選択や決断の確信を得て、行動計画を開始できる。断片的な事前知識間の関係付けを明らかにして、組合せによる利点を新しい視点から行動に活かせる。しかし、小文字での説明は、有用であっても読むのが厄介なため敬遠される。地図が読めない人や読み間違える人には、優しい案内は有用であるが、小文字だと敬遠される。高齢者や視覚障害者は、優しい説明をラジオを聞くように音声で聞けることで、容易に事前情報に現地情報を編み込むこと

ができる。

(iv) 臨場感のあるストーリー：実話や臨場感のある音により、まるで実際のシーンに居るかのよう体感できる。訪問者は信念をより強固にでき、より好い選択も行える。例えば、パレードの実況音を聞けば、フェスティバルに参加したくなる。プレイ中の掛け声を聞き、スター選手を見たりすれば、試合の観戦に行きたくなる。深い感銘を受けると共感を共有する間になれる。共感を得て、新たな行動を起こすことにも発展する。臨場感のある表現は、感情的な情報や評価も伝えることができる。

3.3 視覚による情報提供システム

人間は70%以上の情報を視覚から獲得する。視覚情報だけの情報獲得には、つぎの6つの問題がある。

- (P1) **読み難い小さいフォントの文書**：多くの人は、小さいフォントの文書は読みたくなく、わざわざ読むために画面に近づく行為も行いたくない。とくに、50歳以上の高齢者は小さいフォントを避ける傾向がある。
- (P2) **潜在的知識**：一般的に人々は、視覚情報を見るだけでは、基礎知識無しでは、深い意味や背景にある経験を見つけることはできない。この重要な知識は、聞くことにより容易に得られるべきである。
- (P3) **不十分な知識**：目的達成に必要な十分な情報を、短時間の視覚情報だけで獲得することが困難である。
- (P4) **複雑な操作**：利用者は複雑な操作説明を読まなければならないだけでなく、システム主導の操作を正しく最後まで完了するまで、ここの操作を一つ一つ理解していく必要がある。
- (P5) **騒音弊害**：音声案内は分かりやすく簡単で有効な方法であるが、空間全体の騒音となる。
- (P6) **同じ内容の繰返し**：同じ内容が繰返されるため、周辺住民に嫌気を与え、新鮮な知識・情報を与えることができない。

4. 視覚情報を補完する静穏並列型音情報システム

一般的に人間は、一つの視覚コンテンツを見ているときに、他の視覚コンテンツを見ることはできない。人間は、視覚コンテンツに見入って、その多くの知識の中から自分に有用な知識を抽出する。知識の抽出は、基礎知識の無い人には、多くの時間を費やしても残念ながらできない。音による知識情報システムは、視覚コンテンツに隠れた深い意味や複数視覚コンテンツに跨る意味を分かりやすく説明できるため、視覚情報システムの補完システムとして非常に重要である。しかし、音声情報提供システムは、騒音の

元となるため、公共の場でそんなに多くは見掛られない。本稿では、騒音を発生を回避できる超音波スピーカーによる静音並列型音声情報提供システムを提案する。

4.1 視覚コンテンツを補完する静音並列型音声情報提供システム

静音並列型音声情報提供システムには、主に6個の特徴がある

- (a) **対象コンテンツの隠れた深い意味の補完**：訪問者は視覚コンテンツや複数視覚コンテンツ間に隠れた深い意味を、とくに苦勞することなく容易に聞くことができる。
- (b) **新たな利点の発見**：音声案内システムは、視覚コンテンツ内の重要な部分を示して重要性を分かりやすく説明するため、利用者は新しい利点を発見できる。
- (c) **並行知識選択**：利用者は、耳・頭の方向を少し変えるだけで、好みの案内を並行して流れる複数音声案内の中から選択して、聞くことができる。
- (d) **並行知識獲得**：訪問者は苦勞することなく、同時に複数の案内を聴いて知識を得ることができる。
- (e) **事前知識に新知識を編み込む**：案内システムが断続的に貴重な情報を喧騒な場所でも提供できるので、通行人は容易に多くの新知識を聞き獲得して、事前知識に編み込むことができる。
- (f) **サービス空間の拡大**：雑踏の中で周りの人が視界を遮る状況でも、数十メートル離れたところまで貴重な案内を提供できる。
- (g) **騒音回避**：幅の狭い超音波ビームにより、飛んでいるときは聞こえないし、聞いている人の耳元に小さい音を再生することができる。これにより騒音の発生を回避できる。
- (h) **喧騒な場所を越えて案内**：どんなに喧しい場所があっても、その空間を越えていき、案内を耳元に再生することができる。
- (i) **エコー無し**：利用者は、エコー／音の震え／尾を引く音が無く、耳元再生の音が非常に聞き取りやすい[Miyachi 07]。
- (k) **コンテンツをテキストで容易に編集・更新**するだけで、TTS (Text To Speech) ソフトで読み上げられる [9]。

4.2 静穏型音声サイネージによるアウェアネスとアフォーダンス

静穏型音声サイネージによる案内機能強化について、3.2節の4種類の機能に対して述べる。

- (i) **アテンション (注意) の獲得**：空港やデパートでのチャイムのように、音サインによって、注意を確実に引くことができる。
- (ii) **キーワードやキャッチフレーズによるアウェア**

アネス：人間は一度に一つの視覚情報しか見ることができないが、静穏型音声サイネージシステムは、周囲 360 度の空間に人間が同時に聴くことができる複数の音声サインを提供できる。通行人は、何をしても何を見ても、容易に音声案内を聞くことができる。また、キーワードやキャッチフレーズの響きやリズムが自然に記憶されるため、自然かつ無意識に記憶できる。一つの単語でも「ことだま」と呼ばれる魂が宿り意味を持って一人歩きして、人々が体験した楽しい記憶を蘇えらせる。

また、次のような難しいシーンでも警告やサインを提示できる。

- ディスプレから離れていてキーワードや文字が読めない
- 周りの他の人に、ディスプレイへの視線を遮られている
- 混雑時に他の通行人のバッグなどをぶつけられないように、用心して歩いている

(iii)長い文による説明。訪問者は、混雑して騒がしい空間を移動していても、PS が伝える重要な説明文を、ラジオを聞くように容易に聞き続けることができる。周囲 360 度の空間から PS により伝えられる複数の長い説明は、(ii)の難しいシーンでも聞くことができる。通行人は、様々な方向から同時に聞こえてくる複数の説明の中から、自分の気に入った話題を選択できる。詳細な説明を聞くことにより、訪問者は次の行動のための決心を固めたり、盛り上がりたりできる。視覚情報に隠れた意味を周囲から音・音声によりアフォーダンスとして得られるとき、無意識のうちにその意味に従った行動を、訪問者は取ることができる。同様な背景を持つ人々に、複数の説明を並行して提供することもできる。訪問者はその中からトピックを選んで、楽しみを拡大できる。このように聴覚による知識獲得は容易であるが、一方で長い聴講時間が必要である。そこで、速聴システムが PS や TTS を用いて研究されている[Miyachi 07]。

デジタルサイネージとの煩わしいインタラクションも、音声案内システムは優しく教えてくれる。小さいフォントで書かれた指示や文書を何度も読まなくてもよくなる。音声サイネージシステムは、さらにつぎの困難なシーンにも有効である。

- 歩行者の視界から数十秒でディスプレイが消えてしまう
- 長いテロップの表示がすぐに終わらない
- 短い説明の表示フォントが小さい
- ディスプレイの設置場所を見つけて、情報入手のために近くまで行かなければならない

(iv) 生き生きした臨場感とストーリー：リアルな話と

臨場感のある音は、対象物の価値を著しく高める。リアルな音は強い印象と実際の状況を伝える。たとえば、肉を焼く「ジュー」という音は、肉汁と焼けて美味しくなった肉の表面の状態を伝えるのに十分であり、聞いた人をレストランに向かわせる。同様の経験を持つ人々に成功への行動を開始させる。親しみのあるアクセントでの経験談は懐かしいシーンを思い起こさせる。PS の利用で騒音の発生を回避できるため、少し長い時間の掛かる説明でも聞くことができる

4.3 脳の音処理の最小化

説明を聞く場合、楽に聴くことができるとよい。音は脳で認識するが、認識し易い条件とは何かを検討する。脳では、認識、修正、記憶の処理を行うが、脳の処理能力には限界がある[Baddeley 00]。PS では、たいへん鮮明で反響が少なく、尾を引かない高い指向性の音を発生するため、個々の単語を明確に分離して聞くことができる。超音波スピーカ (PS)の音は耳元で再現されるために、次の5つの特長がある。(1)反響音の回避、(2)音の震え現象の回避、(3)尾を引く音の削減、(4)非日常の音空間(頭内定位音場、耳元音場)、(5)高い指向性。

これらの特長により、利用者は、音による単語認識の確認、修正から開放され、同時に説明内容のコンテキストを記憶するための内的な余裕ができる。この脳内でのリアルタイムの複雑な音認識処理から開放されることにより、利用者はさらの他の作業を行うための内的余裕を持つことができる。並行して、他の説明を聴いたり、入手した知識を事前知識に組み込んだりできる。基礎知識を持っていれば、より容易にこれらの処理ができる。

5. 高指向性音声案内向けの長い説明文のブロッキング法

訪問者は周囲 360 度から同時に複数のコンテンツを聴くことができる。行動の開始や決心するためには、説明の詳細を聞く必要があり、時間が掛かる。本稿では、人間の潜在能力である並行聴取能力を活用して、同時に2つの説明を聴く手法を提案して、合計の説明聴き取り時間の削減を目指す。

人間は、いつも説明のすべての内容を短い時間内に理解できるとは限らない。主な2つの理由は、(1)文章一つの長さ と (2)2つの説明音声の重複である。並行聴取のためのブロッキング法は、マジックナンバー[Miller 56]と脳内での並行処理の両方を考慮する必要がある。

5.1 長文の容易な聴取のブロッキング法

利用者は、各文章を順に理解して、コンテキストを段階的に記憶していき、説明全体を理解する。人間は短期記憶の限界数マジックナンバー：7±2を持っている。長い文の単語数は簡単にマジックナンバーを超えてしまう。長い文の容易な聴取りと長い説明全体の容易な理解のために、聴き取り易い複数の単語の塊（チャンク）の構成要素数を何個にすればよいかを検討した。実験結果から、日本語の長文は7±2個以下の単語で構成されるフレーズに分割することが日本語のネイティブスピーカーに聴き取りやすいことが判明した。容易な聴取りを可能にするブロッキング法(i) - (iii)はつぎのとおりである。

- (i) フレーズ自身が意味を持つ範囲で、7+2個よりなるべく少ない単語数のフレーズを作る
- (ii) フレーズが意味を持つために最低限必要なつぎの単語をフレーズに追加する
- (iii) フレーズを一気に読み上げて、フレーズ間に間をおいて、説明文を読み上げる。

Test 0. マジックナンバーより小さい3種類のフレーズの作成

マジックナンバーの条件を満たす約200文字の3つの説明文：Ea, Eb, Ecを作成した。

Ea：ダイエットにはテニスがお勧めです。ストレス解消につながる爽快な打球感、ゲーム性がある、脳にも刺激があり健康になれます。テニスは適度な有酸素運動なので、脂肪の燃焼を促すため、ダイエットの効果大です。ゆっくりとしたペースで、軽くジョギングをするように足を動かし続け、ひたすらラリーを続けることが、最も効果的なダイエット法と言えます。テニスで輝く美しいダイエットを実現しましょう。

Eb：ウィンブルドン、テニスの四大国際大会の一つ。四大大会中最も古く、120年以上の歴史を持ち、唯一芝生のコートで行われる。試合と練習中は白いウェアとシューズが義務づけられている。開催国イギリスの優勝者は、男子は1936年、女子は1977年を最後に出していない。1995年、松岡修造が日本人男子選手として、当時62年ぶりにベストエイトに進出した。今年の大会で、ロジャー・フェデラーが五連覇を達成した。

Ec：マリア・シャラポワ、2007年世界ランキングは第5位、誕生日は1987年4月18日の二十歳、身長188センチ、体重59キロ、右利き、得意ショットはフォアハンド、バックハンドは両手うち。使用ラケットはプリンス、シューズとウェアはナイキ。17歳

でウィンブルドン制覇、これは史上二位の最年少記録。彼女がボールを打つときの声量は101デシベルで、パトカーなどのサイレンや道路工事などと同等レベル。

文字数の多いフレーズは、カタカナ単語を多く含む場合と、ルールや状況を記述する場合であり、文字数は、15以上31以下であった。実験から、7+2個以下の単語数のフレーズを、説明文全体に対して、容易に人手により作成できることが判明した。フレーズを構成する平均7個の単語は、平均14文字であった。また、各フレーズには、音の響きとリズムがあり、容易に聴き取れ、記憶できた。この結果、長い文章のマジックナンバーに基づくフレーズ分割による音声説明の容易な理解の実現が確認できた。

5.2 長い2つの音声説明を短時間で理解できるための調整手法

通行人は、複数の長い説明を聞くための時間を十分持っていない。その一つの対応策が速聴手法[Miyachi 07]である。本稿では、複数のコンテンツの同時聴取りにおける、容易な興味コンテンツの選択と、長い2つの音声説明の並行聴取りを可能にする並列音声案内システムを提案する。説明音声の重複は、他の説明文の理解を妨げる。本稿では、説明音声の重複を回避する並列音声案内のための調整手法を提案する。基本的な手法は、一方の説明の音声フレーズの間、他方の音声フレーズを聴取りできる同期をとることである(図2参照)。この手法は、4.1節で示したPSによる音声案内システムの特長があるため実現可能になる。すなわち、たいへん明確で、エコーと音の振るえ(空気の流れなどによる)がなく、尾を引く音が無い音が、高い指向性を持ち異なる方向から複数聞こえるという時空間を作れるからである。これにより、個々の単語が明確に分離して聞こえるため、音声の認識における再確認や修正が不要になり、内的余裕が作れる。これにより、説明のコンテキストを同時に記憶でき、複数個の説明を並行して聞くことが可能になる。

■2つの説明の同期手法

(i) 2つの音声説明(E1, E2)に異なる人の声をText To Speech (TTS) systemにより生成する

(ii) 2つの音声説明を重複が無いように同期させる
STj(i) : Ejのi番目のフレーズ。

ETj(i) : Ejのi番目のフレーズの聴取時間。

初期値の設定 : i=1, ST1(1)=ST2(0)=ET2(0)=0

for all i, E1, and E2,

ST1(i) ←ST2(i-1)+ET2(i-1),

ST2(i) ←ST1(i)+ET1(i),

i ← i+1.

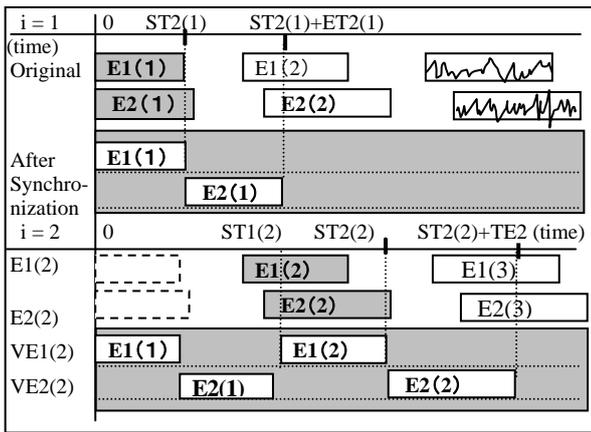


図 5. 2つの音声フレーズの同期化

Ec と Ea の説明の所要時間. Ec の説明時間は、元 : 32.3 秒, 同期化 : 38.9 sec. 約 1.2 倍の時間で Ec と Ea の両方を聞くことができた。

6. 長文説明と並列説明を理解する実験

人間は周辺 360 度から同時にさまざまな音情報を獲得している。鳥の鳴き声、募金の呼び掛け、自動車のクラクション、ストリートダンスの音楽、電車の案内など短時間の音や音声は、人間は苦も無く自然に聞き分けている。観光案内の音声説明も短時間で話せる単語が多く、方角、距離、有名なランドマーク名などがある。

長文による説明には、施設やイベントの歴史と説明、練習方法の順を追った説明、有名選手のエピソードや特徴の説明などがある。実験では、長文の説明に最近人気に戻ってきたテニス関連ビジネスに関係する 3 種類の説明を準備した。

■実験条件

フレーズと説明の長さ: フレーズは (7 + 2) 単語以下の長さ。説明は約 200 文字。

説明: 英国ウィンブルドンのテニスコートとウィンブルドン 5 連覇のフェデラー選手、テニスによるダイエット方法、シャラポア選手の大きな掛け声やブレイの紹介。

コンテンツ: Ea はウィンブルドンの優勝経験者であり、TV 出演も多い Maria Sharapova の紹介である。説明の主な内容は、シャラポアの掛け声がパトカーと同じ 100dB であることや、彼女の様々な経歴やデータが中心で訪問者にも聞きやすい内容である。Eb は、ウィンブルドンの場所と歴史、さらには 2 名のテニス選手、5 連覇のロジャー フェデラーと日本人ベスト 4 の松岡選手の紹介である。Eb は、歴史に関する長い説明が特徴である。訪問者は、説明全体に聞き入ることになる。Ec はテニスを楽しみながらダイ

エットをどのようにするか説明である。多くの人が健康とダイエットに興味があり、具体的な技術、体の動かし方と原理を、順を追って説明する。訪問者は、体の動かし方をよく見て頭の中で模擬して順にそのイメージを記憶する必要がある。

実験 1: パラメトリック・スピーカ (PS) と通常スピーカの比較: 2つの説明の同時聴取り.

場所: 東海大学 2 階建て H 棟の屋上

被験者: 8 人. 20 歳台 7 名、50 歳台 1 名.

コンテンツ: Ea, Ec

スピーカ: (i) PS 2 台,

(ii) 音響スピーカ 2 台.

スピーカの間隔 4m. 被験者とスピーカ間 2.5m.

設置高さ: 1.7 meters (図 6 参照).

TTS システム: スマートボイス [NEC 07]

騒音環境: 約 56 dBA (RION NL-26 使用).

音響スピーカ. 全被験者が両方の説明を理解できなかった。2つの説明が混ざりあって一つの音になり、2つの音空間が一つの音空間になったため聞き取れなくなった。

パラメトリック・スピーカ [Kyouno 04]: 全被験者が両側から説明を聞くことができた。音声は右からの音声は右耳、左からの音声は左耳のそれぞれの耳元で再生されたため、エコーが少なく、両方向から来る音声を聞き分けられた。

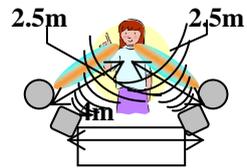


図 6. PS と音響スピーカ

実験 2: 3つの説明が流れる時にその1つの説明を聞く

音響システム: PS を左右両側、前方にラウドスピーカ (図 7 参照). 2種類の妨害音声流れる中で説明を聞く。妨害音声は、通常の音と超音波スピーカによる音の 2 種類。

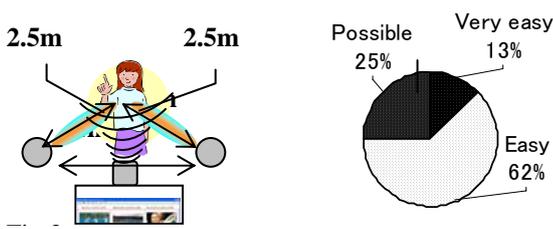
場所: 東海大学 9 号館 4 回の階段上の踊り場

Data: Ea, Eb, and Ec. TTS システムにより 3 種類の男性/女性の声を生成。

騒音環境: 約 47 dBA.

結果. 被験者全員が 3 つの説明のうち 1 つの説明を理解することができた (図 7 参照). 長文による長い説明をチャンクサイズ (7±2) 以下のフレーズに分割することにより、容易に理解できることが確認できた。また、左側と右側から来る音声説明が各々の側の耳元でささやくように聞こえるため、耳打ち音場が左右に生成されて理解しやすくなっていることも判明した。チャンクサイズと音声の高い指向性の相乗効果が確認できた。

二人の被験者は、説明の一部は分かったが説明全体を理解できなかった。この二人はテニスにもダイ



(a)テスト2, 3 の環境 (b) 1つの説明の聴き取り結果
図7. 音声説明の聴き取りやすさの実験

エットにも興味が無く、有名な名前や基礎的単語もよく知らなかった

■被験者へのインタビュー： 7名の被験者は、イヤホンで聞くような音場が左右の耳元にでき、容易に説明を聞くことができた。一人の被験者は音が頭内に聞こえるのが奇妙に感じて、よく聴き取ることができなかった。



図8. スポーツ店: ダイエット, ウィンブルドン, シャラポア

実験 3: 3つの音声説明から同時に2つの音声説明を聴取る

■ 条件がテスト2と同一。視覚情報(静止画、動画)で注意喚起と要点表示を行う(図8参照)。

テスト 3-1 ラウドスピーカーによる視聴

被験者全員が幾つかのキーワードを聞くことができたが、説明全体を聞くことや理解はできなかった。

テスト 3-2 PSによる視聴。

6人の被験者は、同時に聞こえる3つの説明から、2つの説明を同時に聴いて内容が理解できた。前方のラウドスピーカからの音声が必要な説明音声を妨害する状況であったが、2つの説明は、左右両側から左右それぞれの耳元で明瞭な音声で聞けることが確認できた(図6参照)。2つの説明は、エコーが無く、周囲の音場と異なる特別な音場で、音源に近い左右それぞれの耳元での囁きのように再生される。特別な音場での男性/女性の合成音声に対するカクテルパーティ効果により、同時に聴いても区別して聴くことができる。また、長文を聴き易くするブロッキング手法、音声の重複を回避する同期法、エコー/

震え/尾を引く音が無いPSの特長の相乗効果も確認できた。2つの説明聴取の所要時間は、60%に削減できた。また、視聴者は、容易に3つの説明から好みの説明を容易に選択できることも確認できた。

2人の被験者は、1つの説明を理解することはできたが、2つの説明を同時に並行して理解することができなかった。これは、多くの人間は、2つの説明文を並行して聞き理解する潜在的な能力を持っているが、一部には、キーワードを知らず基礎知識が無いため、説明を聞きたくなくなる人もいるからである。実験から、長い説明に興味を持たないために説明を聞くのを止めた被験者が1名確認できた。

案内システムのビームの外側は、説明が聞こえないで静かであった。PSの音が聞こえている所でも、1~2m移動するか、首の向きを変えることで、説明を聞かなくても済むようになる。暗騒音が大きいと、PSの音が聞こえるビームの幅はより限定される。

■被験者の話

1人の被験者は、一つの説明に夢中になると、他の説明を聞くことをいつの間にか忘れてしまう。この被験者は、普段から一つの話だけに集中して聞くために、同時の他の説明を聞くことに慣れにくい。

7. ま と め

訪問地での知識の提供は、訪問地だけで理解できる場合も多く、非常に有効である。準備不十分の訪問者に、視覚情報だけでは伝えられない臨場感や隠れている魅力を、共鳴知、シーン知、環境知、実践知により音により与えることにより、満足度の高い行動支援が行える。聴覚による知識提供により、ディスプレイが見え難い離れた所からでも、注目や共感を、得られるようになり、観光事業や内需の拡大に貢献できる。長い文章の説明文を読むことは厄介であるが、視覚情報に隠された深い意味や複数視覚情報に跨る意味の説明を聞くことは容易である。複数の長い説明を、容易に聞き取れ、並列に効率よく聞取るために、ブロッキング法と複数音声説明の同期法を提案して、有効性を確認できた。人間の認識能力の限界であるマジックナンバーや脳内処理量を考慮して、潜在的な人間の音知識の聴き取り能力を活用する方法である。実験により、2つの同時音声説明を約75%の被験者が並行して聞き取れた。これにより、PSを用いた並行音声案内提供システムの有効性を確認できた。

◇ 参 考 文 献 ◇

[Baddeley 00] Baddeley, A.D. 2000. The episodic buffer: a new component of working memory? Trends in Cognitive Sciences,

4, pp.417-423.

[Rose 06] Bill Rose, Diane Williams, 2006. The OnSpot Digital Advertising Concept, Arbitron. <http://www.simon.com/arbitron/OnSpotDigitalAdvertising-ArbitronReport.pdf>

[ecute 08] ecute Japan. 2008. ecute. <http://www.ecute.co.jp/>

[Franconeri 05] Franconeri, S. L., Hollingworth, A., & Simons, D. J. (2005). Do new objects capture attention? *Psychological Science*, 16, 275-281.

[Fujitsu General 08] Fujitsu General Co. 2008. UBWALL: <http://www.fujitsu-general.com/jp/products/ubwall/index.htm>

[Gibson 66] Gibson J.J. 1966. *The senses considered as a perceptual system*, Boston, Houghton, Mifflin.

[Miller 56] Miller, G. A. 1956. The magical number seven, plus or minus two: Some limits on our capacity for processing information, *Psychological Review*, 63:81-97.

[Kyouno 04] Noboru Kyouno. 2004. *Technology Trends on Parametric Loudspeaker*, JSME.

[NEC 07] NEC. 2007. SmartVoice: http://121ware.com/product/software/smartvoice_4/.

[Miyachi 07] Miyachi, T., Balvig J. J., Jiang Shen Bo, et al. 2007. A fast reading spatial knowledge system by ultrasonic sound beams, Springer, LNAI 4694.