大型映像コミュニケーションとモバイルのための音声技術

梶 克彦（名古屋大学大学院工学研究科）

Sound Technology for Large Video Communication and Mobile Application

Katsuhiko Kaji (Graduate School of Engineering, Nagoya University)

1 はじめに

音声は人同士のインタラクションはもちろんのこと，計算機と人のインタラクションにおいても重要な役割を果たす．本稿では，大型映像コミュニケーションにおいて存在感を高めるための音声再生方式と，スマートフォンによる音声のみを用いたナビゲーションシステムについて述べる．

2 大型映像コミュニケーションシステムにおける音声再生技術

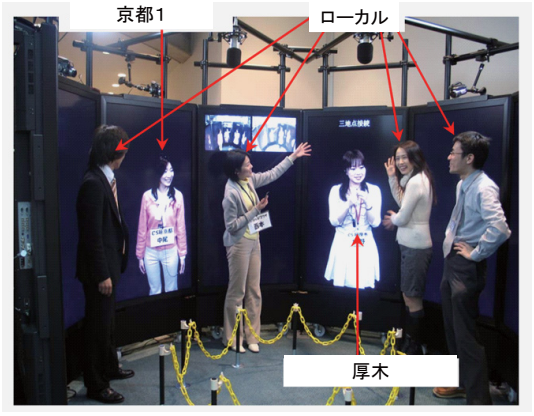

Fig. 1: t-Room による 3 地点の接続

t-Room は，遠隔にいる人同士があたかも同じ部屋にいるかのような感覚（同室感）を得られるよう設計されたシステムである (1)．t-Room は複数の大型ディスプレイとカメラで構成される．それぞれのディスプレイの表面をカメラで撮影し，遠隔 t-Room の対応する箇所のディスプレイにその映像を等身大表示する（図 1）．この構成により，t-Room 内では，視線や指さし等の方向感が保存される．誰が誰を見ているか，何について話をしているかがわかるようになり，自然なコミュニケーションが可能となる．

同室感の実現に際して障壁となったのが，遠隔音声の再生方法である．遠隔の人が声を発したら，その人が声を発しているのだと気づける必要がある．そのためには，音声の発生源である人の映像の方向から，音が聞こえなければならない．

一般的にスピーカは，図 2(a) のように，ディスプレイの両端に配置される．それらのスピーカの音はディスプレイの正面に立つ A,B,C すべての人に届く．この時，認識される音の発生源の位置は A,B,C 全てにおいて異なる．例えば両スピーカから同時に音を発した場合，正面に立つ B は，ちょうどディスプレイの中心から音が聞こえるように感じる．しかし，正面から外れている A や C は，音源の方向を，自身に近いスピーカ（直接音の発生源）の方向と感じる．この効果は先行音効果として有名である．音源となるスピーカが 1 つであり，かつディスプレイに映っている人の口の位置にそのスピーカを配置できれば，この問題は生じない．しかしハードウェアの制限から，ディスプレイ面にスピーカを配置し，かつディスプレイの映像を覆い隠さないようにするのは困難である．

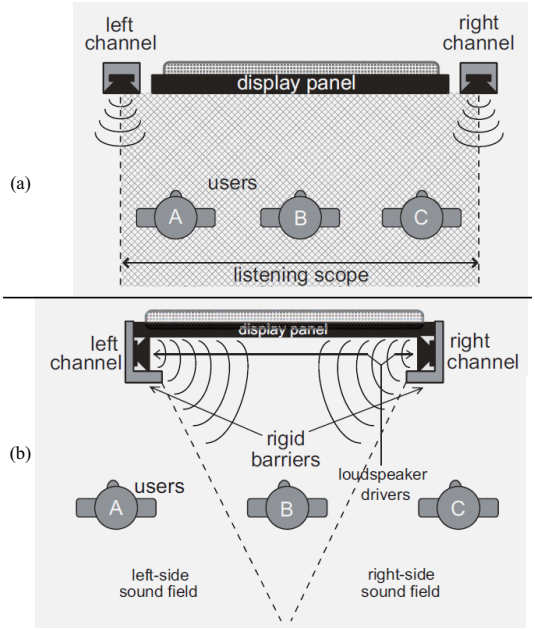

Fig. 2: (a) 一般的な配置のスピーカ，(b) スピーカを内側に向けて配置

この問題を解決するために，仮想的に音源をディスプレイ内の任意の 1 点に見せかける手法を提案している (2)．図 2(b) のように，ディスプレイ側面に，内側を向くようにスピーカを配置する．またスピーカには L 字のバリヤが装着される．L 字バリヤにより，スピーカに近い人にそのスピーカからの音が直接届かないようになっている（図 2(b) のユーザ A や C）．さらに，両スピーカから発せられる音はディスプレイ表面で衝突し，そこからディスプレイ前面の側へ伝わるため，音源の位置が 1 つであるように感じられる．両スピーカからの音の発生タイミングをずらせば音が衝突する位置が変わるため，本手法によってディスプレイの任意の位置で仮想音源を実現できる．t-Room の各ディスプレイにつき，図 3 のようにスピーカが取り付けられる．本手法は，t-Room に限らず大型ディスプレイとスピーカを用いた遠隔コミュニケーションシステムにおいて適用可能である．

3 ナビゲーションにおける音声利用

多くのモバイル用ナビゲーションシステムは，画面（地図）の閲覧を前提としているが，スマートフォンを閲覧しながらの移動は大きな危険を伴う．近年では，駅のホームからの転落や人との衝突などが問題視されている．視覚は周辺環境の確認や危険回避のためにその機能を発揮する必要があるため，それを妨げるシステムはそもそも移動時の使用が好ましくない．

そこで，視覚を妨げずに経路誘導を行うために，音声のみを用いたナビゲーションシステムを提案している (3)．曲がり角や進行方向などを音声のみでユーザに適切に指示するために，

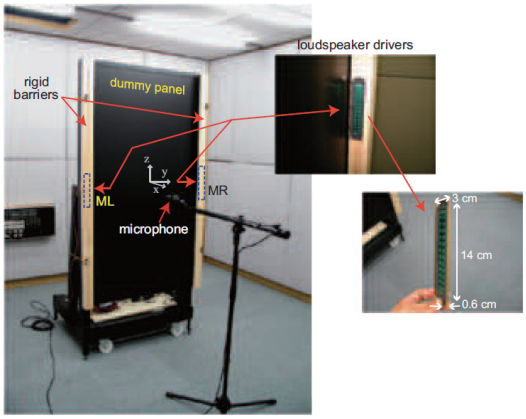

Fig. 3: t-Room スピーカ

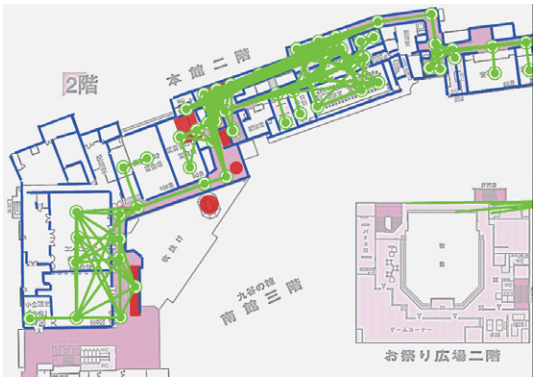

Fig. 5: 空間構造の例．歩行空間ネットワーク（緑色），ランドマーク（赤色），壁（青色）

目印となるような，自動販売機，階段，エレベータといった視認性の高いランドマークを使用する．また，右左折した地点が正しいかが不安になるという問題を解決するために，右左折後に向かうべき目標物も合わせて通知する（図 4）．

Fig. 4: ランドマークに基づくナビゲーション音声テキストの生成

ナビゲーション音声テキストを生成するためには，対象環境内を，人がどのように移動可能であるかを示す歩行空間ネットワーク，目印となりうるランドマーク，各地点からどのランドマークが視認可能かを判断するための壁の情報をあらかじめ登録しておく必要がある（図 5）．歩行空間ネットワークにおける各ノードは，そこを通る際に角を曲がる，階段をあがる，エリアを通り抜けるなどの歩行者の動作が考えられる特別な地点である．この情報を登録するシステムも実装している．登録システムでは，タブレット型端末を用いて，フロアマップの上にネットワーク構造やランドマーク情報をプロットしていくことができる．これらの情報はクラウド上のサーバに送信され，共有されるため，複数人で登録作業を分担できる．

ナビゲーションは以下の手順で行う．ユーザはシステムから現在地と目的地を入力する．入力された現在地と目的地に基づき，歩行空間ネットワーク構造からダイクストラ法で最短移動経路を求める．このとき導かれた経路に含まれるノード N_i について，ノード N_i から次のノード N_{i+1} に移動するための指示と，N_{i+1} に到着した際，N_{i+2} に向かうための動作の指示のための音声テキストを生成する．音声テキスト生成には N_i，N_{i+1}，N_{i+2} の周辺に存在するランドマークを用いる．この際，ランドマークの選択は以下の 3 つのルールに基づいて行われる：(1) 進行方向に存在するランドマークであること，(2) ノードの位置から視認可能であること，(3) 最も視認しやすいランドマークであること．視認可否は，ランドマークの向きや壁情報を用いて判断される．視認しやすさについては，ランドマークの大きさや色などから判断される．ユーザが各ノードに到着した際に，システムがナビゲーション用音声を再生する．音声の生成には一般的な音声合成ソフトを用いる．

このように，使用時の移動を前提としているモバイルアプリケーションにとって，音声は重要なインタフェースとなりうる．視覚を妨げない他のインタラクション手段としては，端末の振動や簡易な物理的ボタンなども考えられるが，伝達できる情報量という点で音声が優っている．

4 おわりに

音声技術は，モバイル・ユビキタス時代のインタラクションを支える基盤技術として今後も重要な役割を担うであろう．本研究分野の今後ますますの発展を期待する．

文 献

(1) Hirata, K., Harada, Y., Takada, et al. t-Room: Next Generation Video Communication System. In *Proceedings of World Telecommunications Congress 2008*. IEEE, Dec 2008. with IEEE Globecom 2008.

(2) Nava, G. P., Shirai, Y., Kaji, K., Matsuda, M., Hirata, K., and Aoyagi, S. Sound Image Localization on Flat Display Panels. In *Advances in Sound Localization*, pages 343–362, 2011.

(3) 渡邉翔太，梶克彦，河口信夫．ランドマークの視認性に基づく歩行者向け音声ナビゲーションの提案．In **マルチメディア，分散，協調とモバイル** *(DICOMO2012)* **シンポジウム**, pages 1897–1903, 2012.