ISS-A-080

t-Room における人物位置に基づく出力選択を伴う音声サーバの開発

吉川　かいり†　　和田　理†　　片桐　滋†　大崎　美穂†
† 同志社大学

1. はじめに

視聴覚メディアの対称性を確保することにより遠隔地との共同作業を支援することを目指して，遠隔コラボレーション支援システム「t-Room」[1]の開発が進められている([2]等)．これまで，その t-Room において，利用者(以下，人物と呼ぶ)が携帯するマイクと壁面部に設置されているスピーカとの間の対応関係が固定されていたために，再生映像と再生音声との間に位置的な不一致が生じるという問題があった．最近，この問題を解決するために，音声出力先スピーカを動的に制御する手法が提案された[2]．本稿では，その動的制御手法を利用して，俯瞰カメラから取得する人物位置に最も近いスピーカに選択的に音声出力を行う t-Room 音声サーバを開発し，その動作の確認を行う．

2. t-Room における音声入出力制御の概要

t-Room における音声の入出力制御に関連するサーバコンピュータの構成を図 1 に示す．

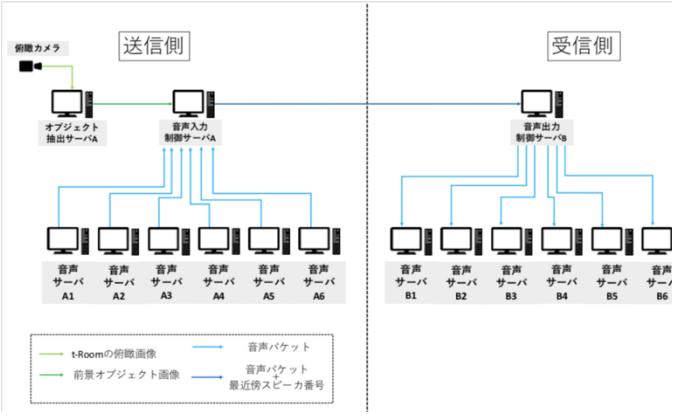

図 1. t-Room の音声入出力を制御するサーバの構成図．

3. 俯瞰カメラによる人物位置推定と選択的音声出力

図 1 に示すように，人物位置の推定には，t-Room の画像入出力制御サーバの 1 つであるオブジェクト抽出サーバ[3]と音声入力制御サーバ(図 1，送信側)との 2 種類のサーバを用いる．オブジェクト抽出サーバは，俯瞰カメラによって取得した画像から人物が存在する領域を示す画像を生成する．音声入力制御サーバはこの画像に対して，スピーカ位置を基準とした領域(スピーカ領域)分割を行い，分割されたスピーカ領域毎に人物領域のピクセル数を計算する．人物領域のピクセル数が最大となるスピーカ領域に対応する(受信側の)スピーカを出力対象スピーカとして選択し，この選択情報を音声出力制御サーバ(図 1，受信側)に伝える．送信された音声データは，音声出力制御サーバを経由して，その傘下の全音声サーバ(図1，受信側)に伝送されるが，実際には共に送られる選択情報に基づいて，出力対象スピーカのみに音信号が出力される(他スピーカ出力は音声サーバ上で抑制される)．ここで，出力対象の選択は，送受信両側における t-Room もそのスピーカ領域も同形であることを利用している．

4. 動作確認実験

受信側の赤印スピーカ(図 2)が選択されるべき送信側 t-Room 内の位置に物体を置き，その物体位置における入力音声として実音声に代えて 440Hz の正弦波を用いて，開発したサーバ群の動作確認を行った．確認の結果，想定した通りに図 2 中の赤印スピーカが選択され，そこからのみ正弦波が出力され，他のスピーカからの出力はないことが確認できた(図 3)．

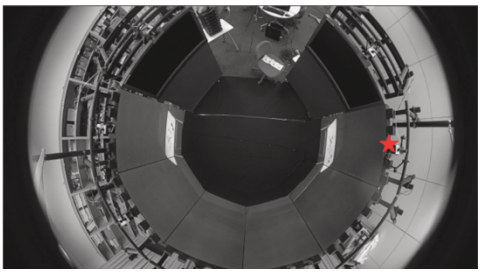

図 2. t-Room の俯瞰カメラの入力画像．

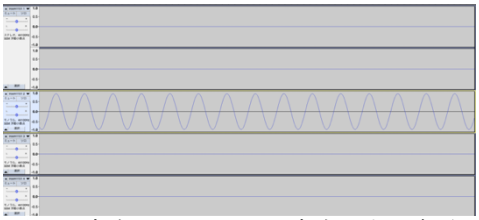

図 3. 音声サーバにおける音声出力の波形．

5. まとめ

t-Room に設置された俯瞰カメラの映像に基づき，人物位置に最も近いスピーカのみに選択的に音声出力を行う音声サーバの開発を行った．今後は複数のスピーカへの選択的音声出力や出力スピーカに対する音圧レベルの制御機能の追加等を行っていく予定である．

参考文献

[1] K. Hirata *et al.*, NTT Technical Review, vol. 4, no. 12, pp. 26-33, 2006.
[2] 東ほか，情処研報，GN-107, pp1-8, 2019.
[3] 和田ほか，情処研報，GN-101, pp1-6, 2017.