

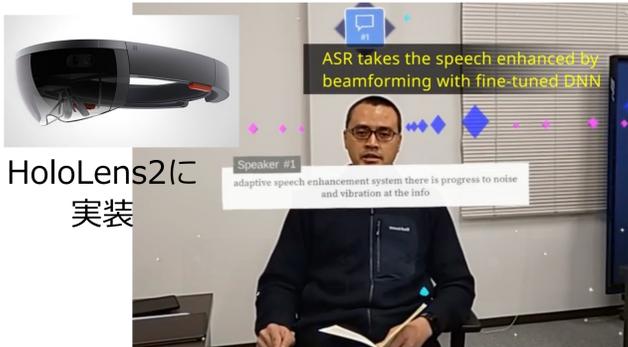
スマートグラスを用いた視聴覚・認識能力の拡張

実環境を想定した高性能・低遅延システムを開発中

日常生活におけるリアルタイム音声強調・認識・翻訳+ARインタフェース

視線/指差しで話者指定・同時発話認識

重複会話の抑制



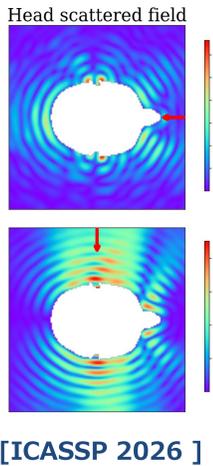
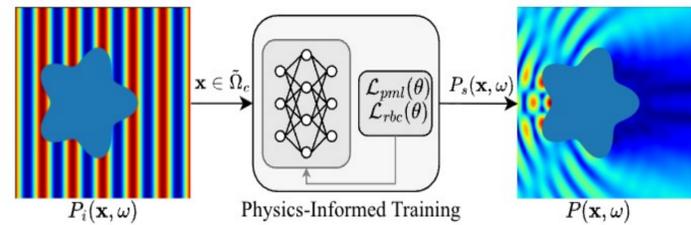
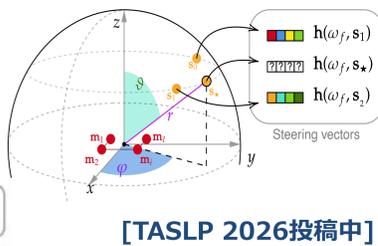
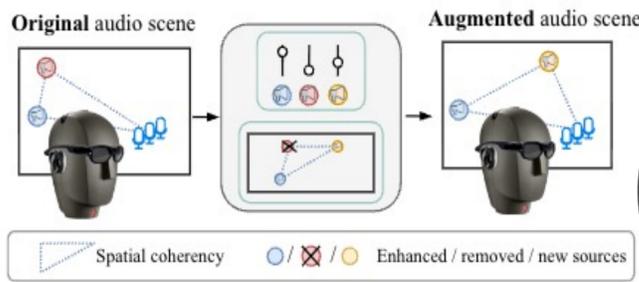
キーコンセプト： 使えば使うほど賢くなる

- 「正解」が得られない実環境で破綻せずに自律的に学習
- オンラインユーザ&環境適応

物理拘束に基づく環境適応型音響レンダリング

- 任意方向に対するステアリングベクトル推定
 - ARグラスユーザの向きに応じた適応的レンダリング
 - 深層カーネル学習 (DKL) + ガウス過程回帰 (GPR)

- 物体周辺の音響散乱場推定
 - 頭部伝達関数推定で重要
 - 物理方程式に基づくDNNの学習



統一的生成モデルに基づく時空間視聴覚場の推論

- 人間の形状・位置追跡と音源分離
 - 屋内に分散配置したマイク・カメラアレイを利用 (位置は既知・同期済み) → スマートグラスとの連携も視野
 - 静止物体に対する時不変モデル → 移動物体に対する時変モデル
- 生成モデルの教師なし学習技術をベースに研究開発
 - 視覚：3D Gaussian splatting (3DGS) + 聴覚：Multichannel nonnegative matrix factorization (MNMF)

