



Silent Communication

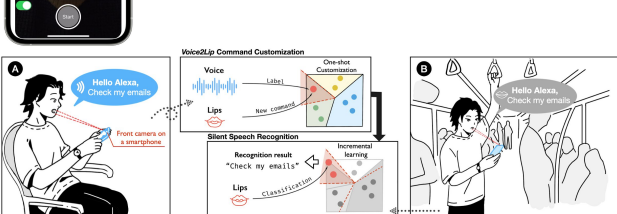
東京大学情報学環教授 ソニーコンピュータサイエンス研究所 フェロー・CSO 暦本純一

概要

AIエージェントとの会話や遠隔コミュニケーションなど、声によるコミュニケーションは次世代HCI (Human-Computer Interaction) の大きな潮流となっています。しかし、公共環境など声を出せない場面も多く、秘匿性の担保もできません。BCIによる発話意図認識は、脳への侵襲が必要で、普通の会話速度での認識速度や精度は達成できていません。本プロジェクトでは、無声あるいは無声に近い微小音声の発話を深層学習により認識する「サイレントコミュニケーション」技術の確立を目指しています。本技術により、いつでもどこでもAIと会話をしながら思考を強化できるthink communicationの世界が到来します。また、遠隔コミュニケーション時の脳同期機構の解明、視線をMLLMに結合し人間の高度技能をAIに理解させる機構の研究に取り組んでいます。

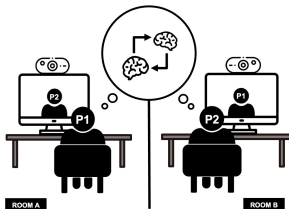
LipLearner: 話者非依存リップリーディング

LipLearnerは、リップリーディングによるサイレントスピーチ技術です。従来のリップリーディングでは話者依存の学習が必要でしたが、本研究では、CLIP (Contrastive Language-Image Pretraining, 対照的言語イメージ事前学習) の手法をリップリーディングに適用し、話者非依存の認識を実現しました。さらに、スマートフォンでも動作するモバイル版により、いつでもどこでも利用できるサイレントスピーチの実現に貢献しました。(ACM CHI2023 best paper)

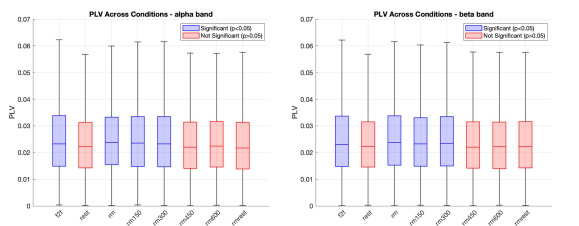


遠隔脳同期機構の解明

共同作業を行う人々の間で脳波が同期することは知られていましたが、遠隔環境でも脳同期が発生するかは未知でした。本研究では通信遅延を制御した遠隔環境での脳同期計測により、450ms以下の遅延環境では共同作業により脳同期が発生することを確認しました。このことから、脳同期の発生機序として、外部要因ではなく脳間の相互作用によることを明確にしました。

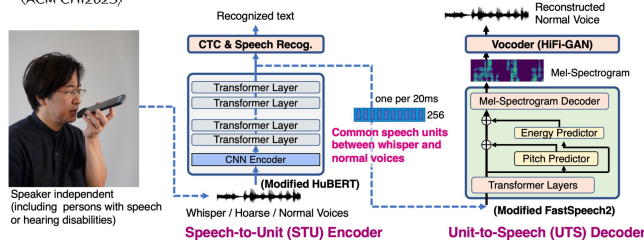


この知見は、遠隔環境での共同作業品質の指標などへ展開が期待されます。(Augmented Humans 2025)



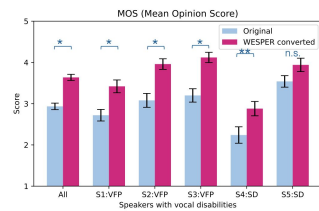
WESPER: 実時間囁き声変換

微小音声による囁き声も、サイレントコミュニケーションの手段として有望です。本研究では、大量の囁き声を事前学習(pretraining)することで、特定の言語や喋り方に依存しないspeech unitを抽出することで、微小音声→通常音声の変換を可能にしています。(ACM CHI2023)



発声困難者の発話支援

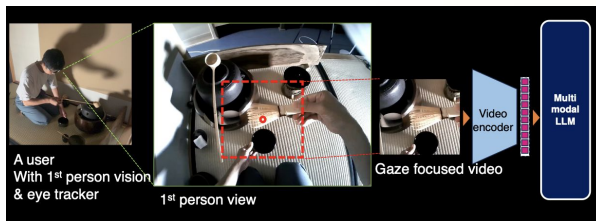
AIによる音声変換技術を活用し発声困難者の発話支援研究を行いました。声帯摘出、声帯ポリープ、ジストニア、聴覚障害など、種々の要因で通常の発声が困難な方の発声からの通常声の変換を行うことで、有意に音質が改善しました。Voice cloning 技術と融合し、各人の本来の声を取り戻す技術を目指しています。(ACM CHI2023)



VFP: vocal fold polyp (声帯ポリープ)
SD: spasmodic dysphonia (痙攣性発声障害)

GazeLLM: 視線を利用するマルチモーダルLLM

一人称視点映像をMLLM (Multimodal LLM) に直接理解させ、人間の技能をAIを介して他者に伝達する機構を構築しています。視線情報を援用することで、技能者の注意をLLMに反映させることができ、行動理解精度が向上することが確認できました。(Augmented Humans 2025)



今後の展望

HCI (Human-Computer Interaction)の今後の大きな潮流は人間とAIの融合です。本プロジェクトで推進しているサイレントコミュニケーションは、AIといつでもどこでも会話のできる世界を構築し、人間の思考が自然にAIによって強化される世界を目指しています。



蘇子雄 (Zixiong Su)
東京大学情報学環博士課程

頼欣好 (Sinyu Lai)
東京大学学際情報学府修士課程

張清 (Qing Zhang, Ph.D.)
東京大学情報学環特任助教