先端科学技術研究科　修士論文要旨

所属研究室 （主指導教員）	サイバネティクス・リアリティ工学 （清川　清　（教授））		
学籍番号	2411126	提出日	令和 8年 1月 19日
学生氏名	迫田　正太		
論文題目	キャラクタに適した話者性を有するマルチモーダルテキスト音声合成		

要旨

深層学習によりTTSは高品質化し，話者性を条件付けて自然な音声を生成できるようになった．その拡張として，顔画像から人物に適した話者性を推定して合成するFace-to-Speech（F2S）が提案され，参照音声に依存しないアニメ調キャラクタへの音声付与にも応用されている．しかし，顔画像のみの条件付けは視覚手掛かりに偏り，性格・背景・役割といった非視覚的個性の反映や制作意図に沿った調整が難しい．また，顔画像から話者埋め込みへの回帰では，外見情報だけでは人物像が一意に定まらないため，推定埋め込みが代表的な領域へ収束しやすく，入力差が音声差として現れにくい．そこで本研究は，キャラクタ設定を自然言語で与えるPersona-Promptを導入し，顔画像と設定文の二つのモダリティから話者埋め込みを推定・統合する枠組みを提案する．設定文を編集可能な入力として扱うことで，外見に基づく直感的な方向付けに加え，人物像に基づく声の印象制御を同一の枠組みで行えるようにする．ユーザスタディの結果，マルチモーダル入力は単独入力に比べて生成音声の自然性とキャラクタとの親和性をともに向上させ，話者の多様性を改善する傾向を確認した．