



音声合成を用いた 音声認識のためのデータ生成

名古屋工業大学 工学専攻
情報工学系プログラム
助教 上乃 聖

本研究の特徴を一言で言うと、、、

音声認識と音声合成の統合による さらなる高精度な音声認識

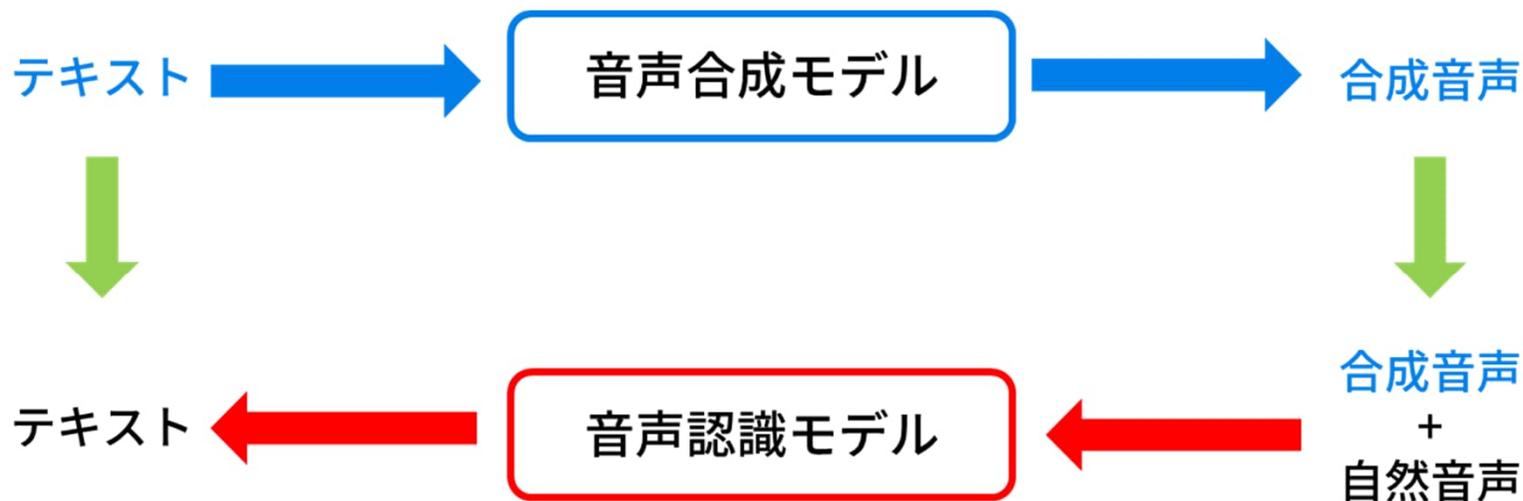
を、研究・開発している。

社会背景と技術的課題

- **音声認識技術の普及**
 - 音声認識: 音声をテキスト化する技術
 - 音声を用いた様々なサービスの基礎として使用
- **専門用語の認識**
 - 一般的な単語の認識は良い性能に
 - 出現場面の少ない専門用語などは認識が難しい

本技術の特徴

- 音声合成により任意のテキストから音声を生成、音声認識で学習することで任意の単語を学習可能
 - 音声認識自体の枠組みは変更なしに学習・使用可能
- テキストデータのみでのデータは比較的準備が容易
 - 本技術はテキストデータのみから音声認識の学習が可能

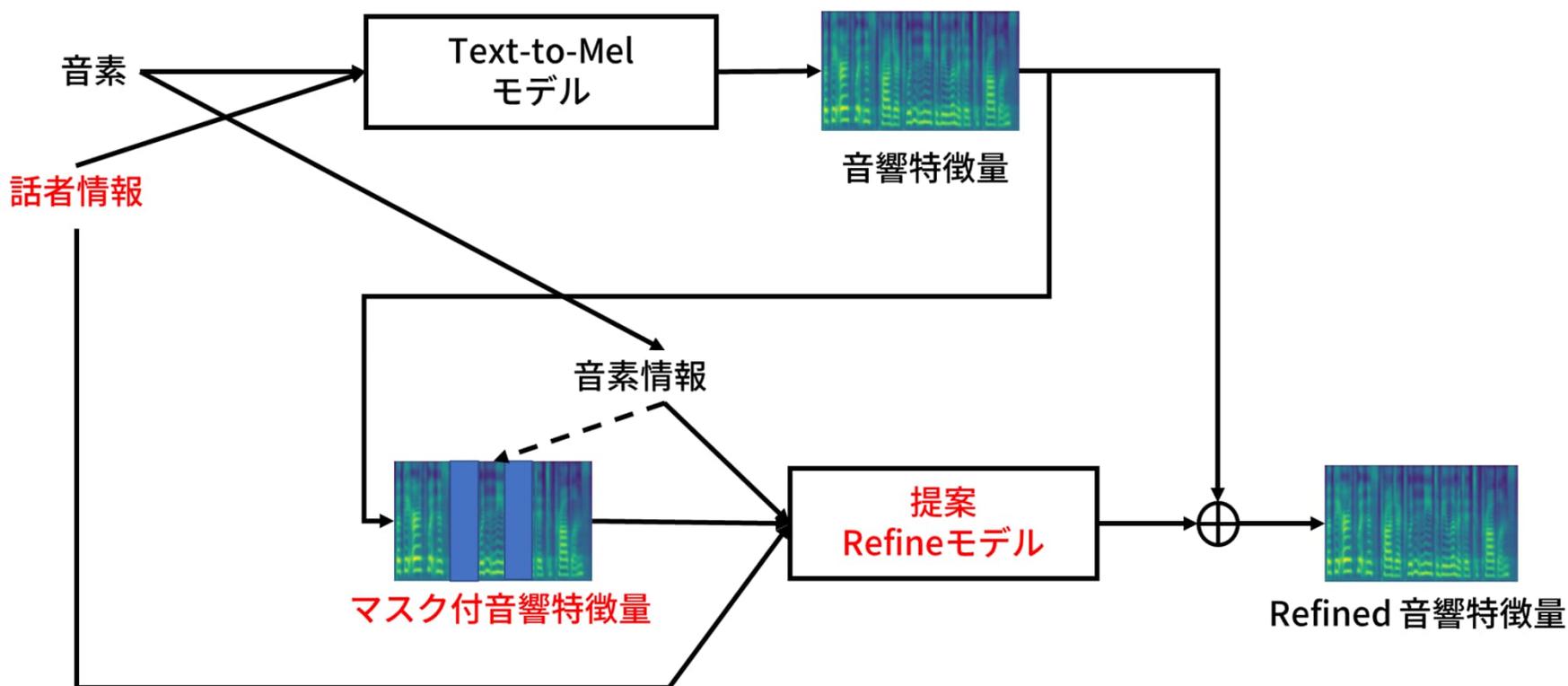


具体的な取り組み

- より効率的なデータ拡張の枠組みの研究
 - 質の向上: 本物の人間の音声に近づけていく
 - 人間にとってよい音声合成と音声認識に取って
良い音声合成は一緒?

具体的な取り組み

- 一度音声を作成し、さらに強調するネットワークを提案
 - 人の声に関わる音素（発話内容）と話者の情報を追加することでデータ拡張の性能の改善を確認



求める連携先・メッセージ

- **本枠組みは非常に幅広い利用範囲があります**
 - 大規模なモデルから小規模なモデルまでどちらにも適用できます
 - タスクに特化した音声認識の構築に役立つ可能性があります
- **ほかにも音声情報処理に関する研究をしています**
 - 幅広くデータ生成に関する研究をしています
 - Speaker diarization、音声変換など
 - 対話などの応用にも取り組んでおります
 - 実サービスのための豊かな音声合成・ロバストな音声認識など

上記の興味があればお問い合わせください

本技術に関する情報

試作品の状況

未定

※提供の際は諸手続が必要となるため、下記問合せ先までご連絡願います。

研究フェーズ



文献・特許の情報

- Sei Ueno *et al.*, Refining Synthesized Speech Using Speaker Information and Phone Masking for Data Augmentation of Speech Recognition, IEEE/ACM Transactions on Audio, Speech, and Language Processing, Vol.32, pp.3924—3933 (2024)
- Sei Ueno *et al.*, Multi-setting acoustic feature training for data augmentation of speech recognition, Acoustical Science and Technology, Vol. 45, pp.195—203 (2024)

【お問合せ】

名古屋工業大学 産学官金連携機構

〒466-8555 名古屋市昭和区御器所町字木市29番

TEL:052-735-5627

E-mail: nitfair@adm.nitech.ac.jp

URL: <https://technofair.web.nitech.ac.jp/>