話者照合コーパス JTubeSpeech-ASV における言語と性別による影響調査

平山 絵理† 塩田 さやか†
† 東京都立大学システムデザイン学部情報科学科

1. はじめに

近年，音声を用いた生体認証技術である話者照合の実用化が期待されており，研究が活発に行われている．実用化を前提とした実データで構成されたデータベースの公開も注目されており，その一環として日本語話者が主となる話者照合用コーパス JTubeSpeech-ASV[1]が公開された．本論文では JTubeSpeech-ASV における言語や性別が性能に与える影響について調査を行った結果を報告する．

2. JTubeSpeech-ASV

JTubeSpeech-ASV とは，YouTube 動画から抽出された音声データから構築される日本語を主とした音声コーパス JTubeSpeech からさらに話者照合用として整備を行ったものである．学習データには日本語，英語，中国語，韓国語など複数の言語が含まれており，評価データは日本語音声のみで構成されている．これまでに全体の性能評価については報告されているが，言語や性別に偏りがあることからそれらの影響についての調査が不十分となっている．

3. 話者照合実験

3.1 実験条件

JTubeSpeech-ASV の学習および評価時の言語と性別の偏りによる影響がどの程度あるのかを評価するため，オリジナルのデータベースを次のように分けて実験を行った．学習データセットとして，JTubeSpeech-ASV の全データを用いたオリジナル（1,792 話者，107,214 発話），日本語話者のみ（979 話者，46,562 発話），オリジナルから男性のみを選んだもの（1,191 話者，78,078 発話），女性のみを選んだもの（293 話者，10,929 発話）の 4 種類を用意した．テストデータのセットとしては，JTubeSpeech-ASV のオリジナル（92 話者，20,976 発話），オリジナルから男性のみを選んだもの（54 話者，7,452 発話），女性のみを選んだもの（30 話者，2,430 発話）の 3 種類を用意した．さらに各テストデータセットには発話長が 5 秒以下（Core），2 秒以下（Short）の 2 種類が用意されている．性別のアノテーションについては人手で行った．

話者照合モデルには VoxCeleb trainer[2]で公開されている ResNetSE34V2 を用いた．評価指標は等価エラー率（Equal Error Rate; EER）を用いた．モデル学習は 500 エポック行い，各条件で EER が最小となるものを結果として示した．

表 1 各条件における EER (%)

学習データ	テストデータ	Core	Short
オリジナル	オリジナル	4.386	5.263
オリジナル	男性	2.174	4.412
オリジナル	女性	8.642	7.500
日本語	オリジナル	4.825	5.817
日本語	男性	3.623	5.147
日本語	女性	8.685	8.750
男性	男性	2.174	4.412
女性	女性	12.345	12.871

3.2 実験結果

表 1 に各学習データセットに対して各テストセットで評価したときの EER を示す．はじめに，言語による影響について述べる．学習データがオリジナルおよび日本語の実験において，学習データが日本語の場合の方がどのテストデータでも EER が低精度となった．これはオリジナルと比較して日本語のみではデータ量が大幅に減るため，データの純度よりもデータ量不足の影響が大きく表れたためだと考えられる．次に，性別による影響について述べる．学習データがオリジナルおよび日本語の実験において，テストデータが男性の場合 EER が減少し，女性の場合大幅に増加した．学習データは男性が大きな割合を占めているため，性別による影響が大きいと言える．学習データが男性および女性の場合においては，上記実験と比べ男性では結果に大差はなく，女性では大幅に低下した．オリジナル学習データの女性話者割合が 16%と少ないことから，全体のデータ量としてはある程度確保されている場合でもデータに偏りがある場合は性別差が顕著に現れることが確認できた．

4. まとめ

話者照合コーパス JTubeSpeech-ASV における性別による照合精度への影響が確認できた．今後はデータの偏りに影響を受けない学習方法やモデル化について検討する予定である．

参考文献

[1] 塩田ら，”JTubeSpeech-ASV: YouTube から構築された話者照合のための日本語を主とした音声コーパス”，情報処理学会研究報告 Vol.2023-MUS-137,No.4,pp. 1-4, 2023.
[2] https://github.com/clovaai/voxceleb_trainer