朗読音声合成におけるポーズ長分布の多様性を吸収するための標準化の効果

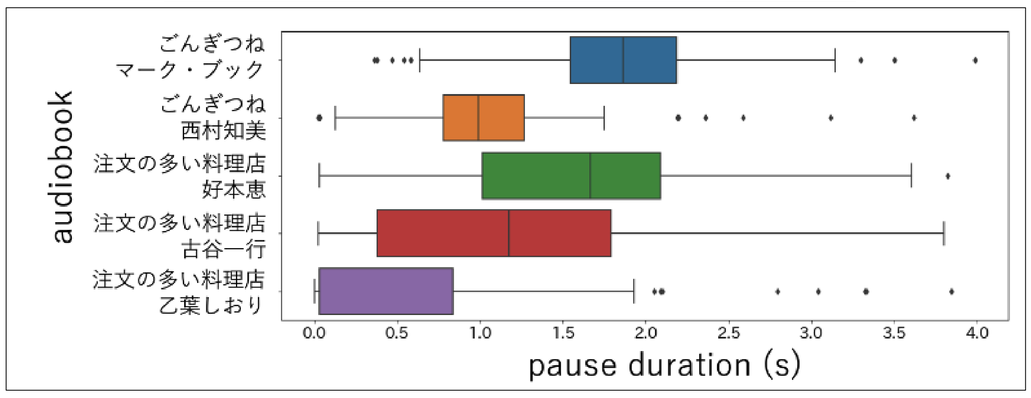
竹下 隼司　松崎 拓也
東京理科大学 理学部第一部 応用数学科

研究背景・目的

背景

自然な朗読音声合成には、正確なポーズ予測が重要
ポーズ位置/長さには朗読者や作品ごとに異なりあり
ポーズ長の標準化で、異なりを吸収できるのではないか？

朗読作品ごとの文間ポーズ長の分布



目的

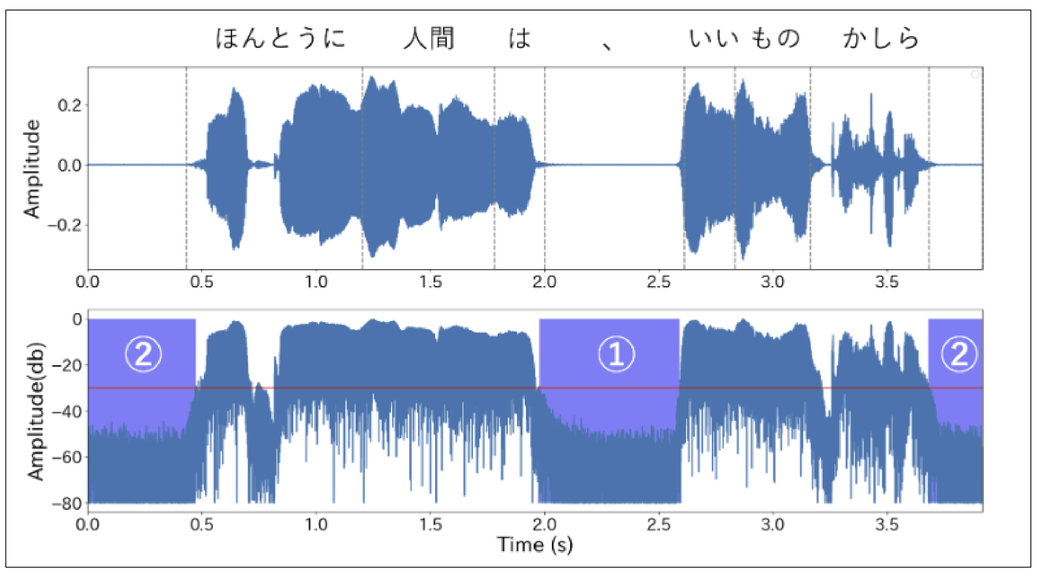
朗読テキストからポーズ位置とポーズ長を正確に予測する

提案手法

ポーズの認定

1. Juliusによる音素アライメントから形態素アライメントを作成
2. 音声波形をデシベル変換し、ポーズ区間を閾値で抽出
3. 抽出されたポーズを、①文中ポーズと②文間ポーズに分類

音声波形とデシベル変換後のポーズ区間抽出結果



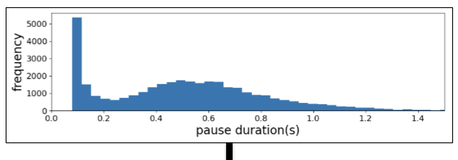
標準化

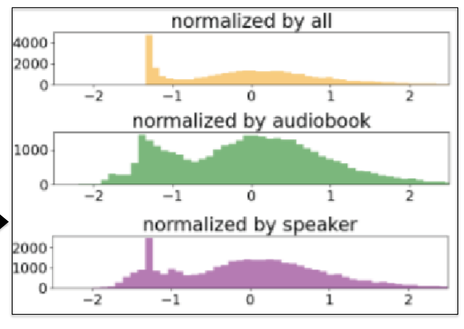
ポーズ長を、話者や朗読作品等ごとに平均と標準偏差で標準化

埋め込み

BERTの最終隠れ層に、話者や朗読作品等の埋め込みを加算

文中ポーズ長分布を標準化





話者や朗読作品ごとに標準化

BERTによるポーズ予測

- モデル構造: BERTまたはBERT+BiLSTMを使用
- タスク: 各形態素の後にポーズがあるかどうか（ポーズ位置）と、ポーズがある場合の長さ（ポーズ長）を予測

実験

実験設定:

日本語多話者朗読作品コーパス（J-MAC）中の全発話に対し、以下の7通りのグループ化を行い、
①グループごとに訓練データを標準化した場合と
②グループの埋め込みをモデルへ追加した場合とで、
文中と文間ポーズの予測精度を比較した

- none　　：標準化なし/埋め込みなし
- all　　：全体に対して標準化/埋め込み
- audiobook　：朗読作品ごとに標準化/埋め込み
- narrative　：ナレーションかどうかごとに標準化/埋め込み
- audiobook-narrative：「朗読作品ごと, かつナレーションかどうか」ごとに標準化/埋め込み
- speaker　：朗読者ごとに標準化/埋め込み
- book　　：文章作品ごとに標準化/埋め込み

結果:

文中ポーズ位置の分類精度

F1-Score	BERT +標準化	BERT +BiLSTM +標準化	BERT +埋め込み	BERT +BiLSTM +埋め込み
none	0.8374	0.8336	0.8365	0.8400
all			0.8365	0.8400
audiobook			0.8334	0.8245
narrative			0.8389	0.8351
audiobook-narrative			0.8341	**0.8416**
speaker			0.8221	0.8384
book			0.8340	0.8404

文中ポーズ長の回帰精度（単位: 秒）

RMSE (単位: 秒)	BERT +標準化	BERT +BiLSTM +標準化	BERT +埋め込み	BERT +BiLSTM +埋め込み
none	0.1522	0.1518	0.1515	0.1541
all	0.1201	0.1203	0.1515	0.1541
audiobook	**0.1128**	0.1133	0.1550	0.1430
narrative	0.1199	0.1202	0.1546	0.1538
audiobook-narrative	0.1129	**0.1128**	0.1558	0.1446
speaker	0.1139	0.1146	0.1537	0.1440
book	0.1186	0.1189	0.1534	0.1509

文間ポーズ長の回帰精度（単位: 秒）

RMSE (単位: 秒)	BERT +標準化	BERT +BiLSTM +標準化	BERT +埋め込み	BERT +BiLSTM +埋め込み
none	0.7179	0.5856	0.7389	0.5824
audiobook	0.6323	**0.4889**	0.6470	0.4898

考察

グループ化の影響

朗読作品によるグループ化が精度が最も高い
→ 標準化後の分布がより正規分布に近く、モデルが学習しやすい分布形状

標準化と埋め込みの比較

回帰精度は標準化が比較的高いが、要因不明
標準化は、埋め込みと比べて学習後パラメータの調整が容易であるが、話者などによるポーズ位置の差異は吸収できない

結論

・ポーズ長の回帰タスクでは標準化が埋め込みより精度が高い
・ポーズ位置の分類では、埋め込みとBiLSTMの組み合わせが最も精度が高く、全体的に朗読作品ごとのグループ化が有効

連絡先：竹下隼司(takeshun1619@gmail.com)、松崎拓也(matuzaki@rs.tus.ac.jp)