



©中山 佐保子¹, ド チュオン¹, サクティ サクリアニ^{1,2}, 中村 哲^{1,2}

¹奈良先端科学技術大学院大学 (NAIST)

²理化学研究所 革新知能統合研究センター (AIP)

{nakayama.sahoko.nq1,do,truong.dj3,ssakti,s-nakamura}@is.naist.jp

コードスイッチング(CS)と音声認識

コードスイッチング(CS)とは、バイリンガル話者が会話の中で二つ以上の言語を混ぜること

文中CS

- [単語レベルCS]: 確率を与えたので、その確率を使って recounting します。 [From NAIST lecture]
- [フレーズレベルCS]: ウルトラマンティガなんだけど悪いティガだから、so what's his name? [Nakamura+2005]

文間CS

ああ、そうだってね。 On the honeymoon, they bought this. [Nakamura+2005]

※本研究は、あらゆる会話を認識することを目指しているため、以下についてもCSの範疇で扱うこととする。

1. 外来語 = 単語レベルCS? 中間言語を使った時のメリットに何かがあるか? [From NAIST lecture]
2. 引用 = フレーズレベルCS? わたしねthat's not the pointとか言われたもんね。 [Fujimura+2013]

日本でもバイリンガルが増加してCSが観察されるようになり [Nakamura+2005], 日英CS音声認識が必要

これまでのCS音声認識

- ① 言語識別を用いて言語ごとに認識 [Mabokerla+2014]
文中CSや言語間で似た単語の場合は特に言語識別が難しく、識別のエラーが認識に伝播してしまう。
- ② 言語モデル, 音響モデル(GMM or DNNベース)をそれぞれバイリンガル学習 [Toshniwal+2018]
音響モデル, 言語モデル, 発音辞書が別々に最適化され、それぞれのエラーが伝播してしまう
- ③ バイリンガル学習したEnd-to-end DNNモデル [Yilmaz+2016] [Seki+2018]
良い性能がみられ、現在の主流となっている。

DNN認識モデルには大規模なデータが必要だが、これまで構築されたCSコーパスは日英以外の言語の組み合わせが多い ←中国語と英語[Vu+2012], フリジア語とオランダ語[Yilmaz+2016] など(日英は、言語間で話者が異なる文間CSのみ[Seki+2018])

- **日英CSに焦点を当て、日英CS音声認識のための日英CS音声データを構築する**
- **バイリンガル音声合成を用い、時間とコストがかからない大規模なデータ構築を行う**

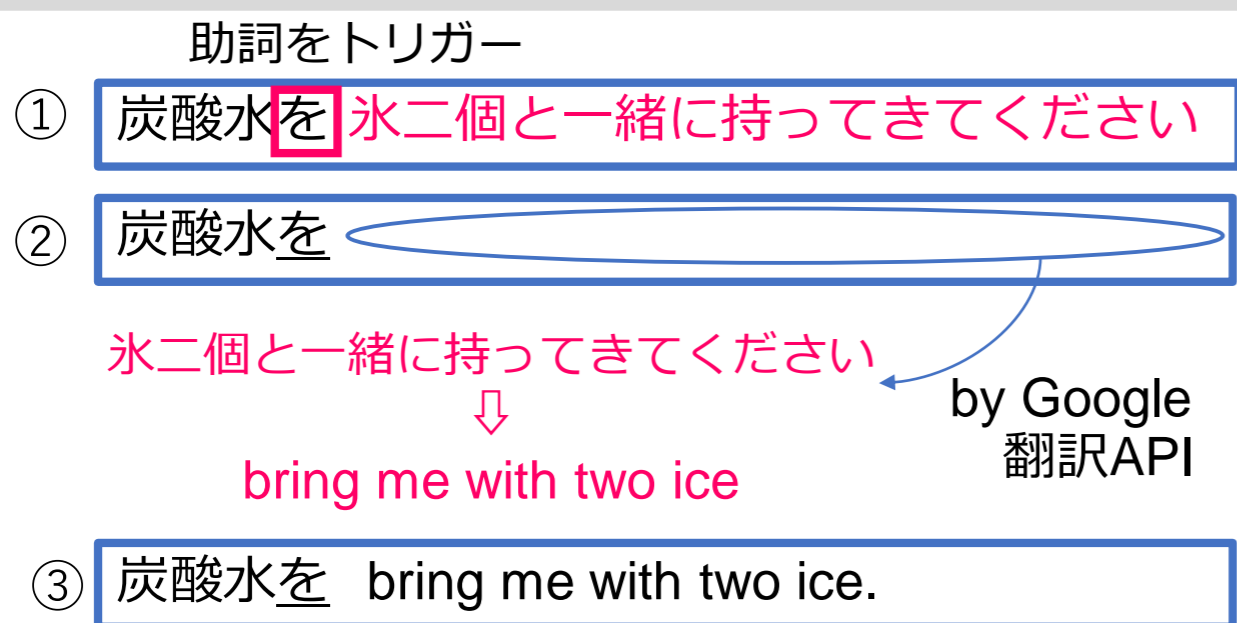
バイリンガル音声合成を用いたCS音声データ構築

バイリンガル音声合成

- ✓ HMM音声合成ツールキット(HTS)[Zen+2007]を用いたオリジナルモデル
- ✓ 強調音声翻訳[Truong+2014]のために収集されたバイリンガルBTEC音声データ(一人のバイリンガル話者による日英各960発話)を学習に使用

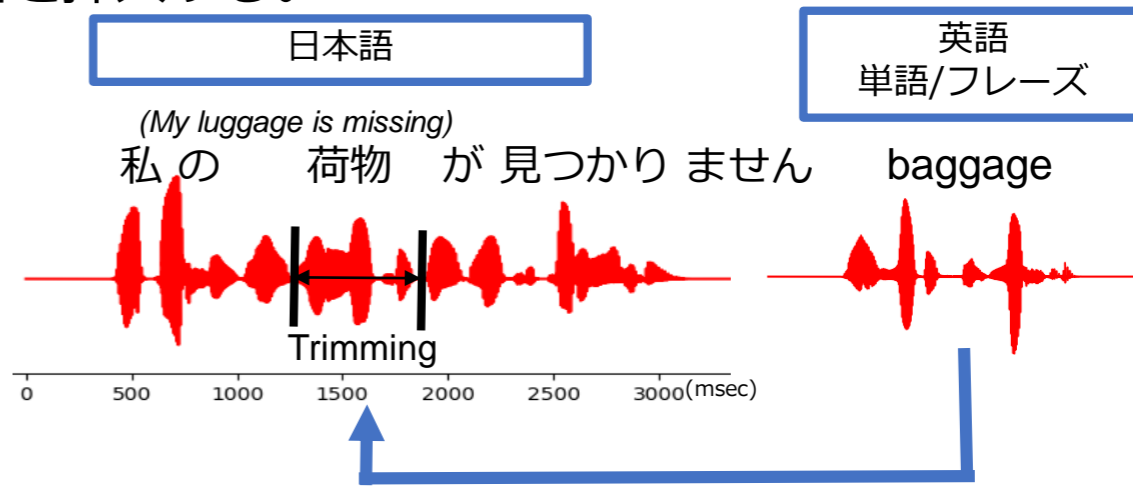
BTECを用いてテキストデータ構築後に、上記のバイリンガル音声合成を用いて音声データ構築を行う

テキストデータ構築



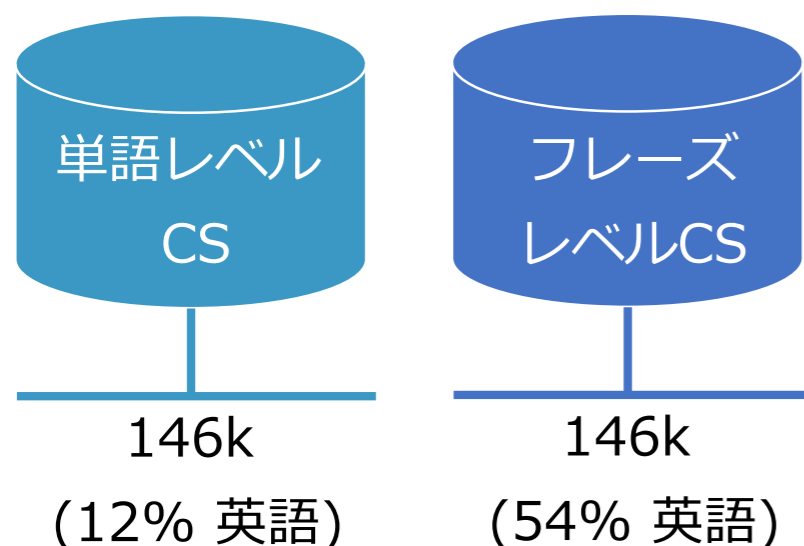
音声データ構築

日本語文全体を合成した後に、音声の時間情報を元に英語を挿入する。

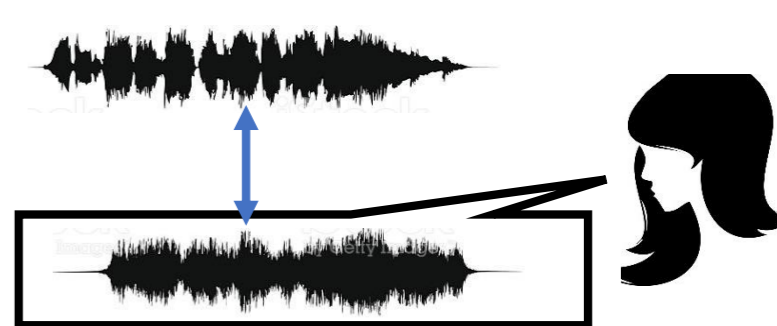


結果と今後の課題

単語レベル, フレーズレベル
各146kのCSが完成



- ✓ 音声の質を高める
 - ✓ 話者数を増やす
 - ✓ 自然な音声との比較
 - ✓ 結合部分の自然な流れへの対応



- ✓ 音声認識システムに実際に用いることでデータの妥当性を確認