



[E4-3]

漸進的な 音声認識・機械翻訳・テキスト音声合成 に基づく 音声から音声への同時翻訳

須藤 克仁, Sashi Novitasari, 帖佐克己, 柳田智也, 二又航介, Sakriani Sakti, 中村哲

奈良先端科学技術大学院大学

2020/03/18

翻訳と通訳、逐次通訳と同時通訳

翻訳

- 静的
- 書き言葉が多い
- スタイルガイド



通訳

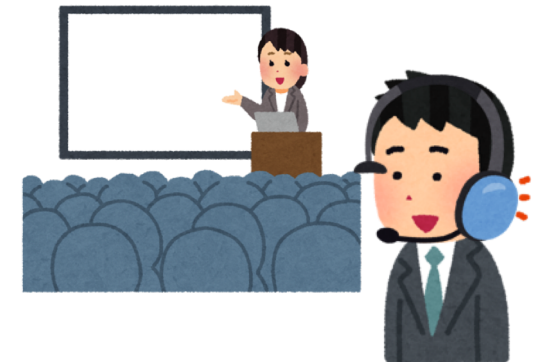
逐次通訳

- 動的？
- 話し言葉
- 時間制約・要約



同時通訳

- 動的
- 話し言葉
- 厳しい時間制約



同時通訳における遅延

(1) The relief workers (2) say (3) they don't have (4) enough food, water, shelter, and medical supplies (5) to deal with (6) the gigantic wave of refugees (7) who are ransacking the countryside (8) in search of the basics (9) to stay alive.

出典: 水野 的 (Akira Mizuno), 『同時通訳の理論：認知的制約と訳出方略』

通常の翻訳

(1) The relief workers (2) **say** (3) they don't have (4) enough food, water, shelter, and medical supplies (5) to deal with (6) the gigantic wave of refugees (7) **who are ransacking the countryside** (8) **in search of the basics** (9) **to stay alive.**

(1) 救援担当者は (9) **生きるための** (8) **食料を求めて** (7) **村を荒らし回っている** (6) 大量の難民たちの世話をするための (4) 十分な食料や水, 宿泊施設, 医薬品が (3) **無いと** (2) **言っています**

「順送りの訳」

(1) The relief workers (2) **say** (3) they don't have (4) enough food, water, shelter, and medical supplies (5) to deal with (6) the gigantic wave of refugees (7) **who are ransacking the countryside** (8) **in search of the basics** (9) **to stay alive.**

(1) 救援担当者たちの (2) 話では (4) 食料, 水, 宿泊施設, 医薬品が (3) 足りず (6) 大量の難民たちの (5) 世話ができない とのことです.

(7) 難民たちは今村々を荒らし回って, (9) 生きるための (8) 食料を求めているのです.

遅延削減に向けた戦略

1. 部分訳が可能になった時点で訳出を開始する
2. 訳出の順序を入力に近いものに変更する
3. 訳出を簡明 (concise) にする
4. 入力を (participation)

現時点では未だ「翻訳」

研究の目標

- 音声言語処理技術に基づく同時通訳
 - 音声認識・機械翻訳・テキスト音声合成の統合
 - 漸進的、かつ実時間動作が必要

漸進的音声認識

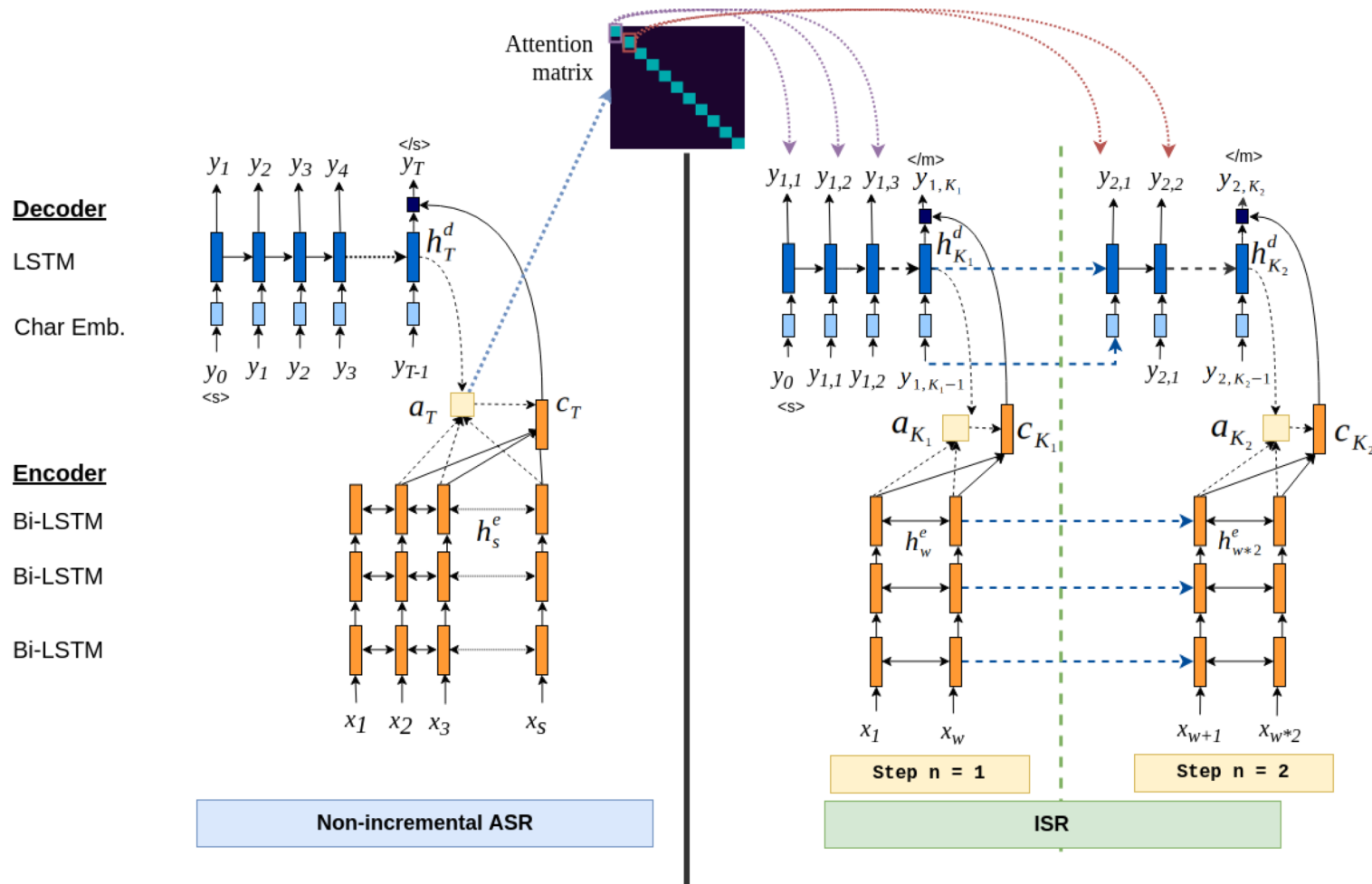
Novitasari, Tjandra, Sakti, and Nakamura: Sequence-to-Sequence Learning via Attention Transfer for Incremental Speech Recognition, Proc. Interspeech 2019, pp. 3835-3839 (2019)

(E4-1: Neural Incremental Speech Recognition Through Attention Transfer)

音声認識と漸進的処理

- 音声認識における文脈情報
 - タスク自体はmonotoneで並べ替えはない
 - 言語モデル観点での文脈は必要
- 漸進的処理を実現する方式
 - HMM: 長距離/後方文脈を考慮できない → 自明
 - End-to-End: 入力全体を注視機構で参照可能
 - 処理単位を分割して参照範囲を限定
 - どう注視範囲を限定させるか？

提案手法：注視転移による漸進的音声認識



- 教師モデルは全体を参照可能
- 生徒モデルは局所情報のみ利用
- 教師モデルの注視を最大限真似る学習

漸進的機械翻訳

帖佐, 須藤, 中村: 英日同時翻訳のためのConnectionist Temporal Classificationを用いたニューラル機械翻訳, 情報処理学会研究報告 2019-NL-241 (2019)

遅延と流暢さのトレードオフ

- 遅延を短くすれば流暢さは一般に低下

I'm going to take you
on a journey
in the next 18 minutes.

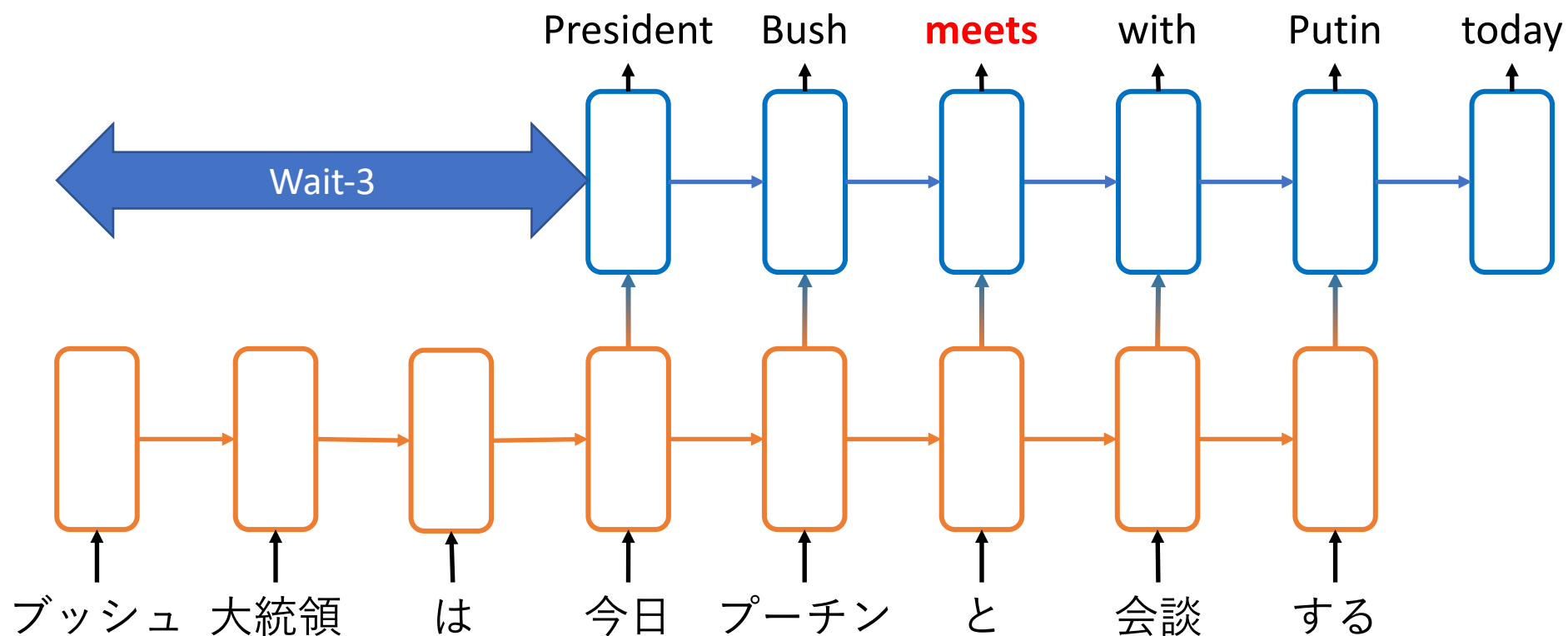
皆様をお連れします
旅へ

今から18分後です

今から18分で
皆様を旅へお連れします

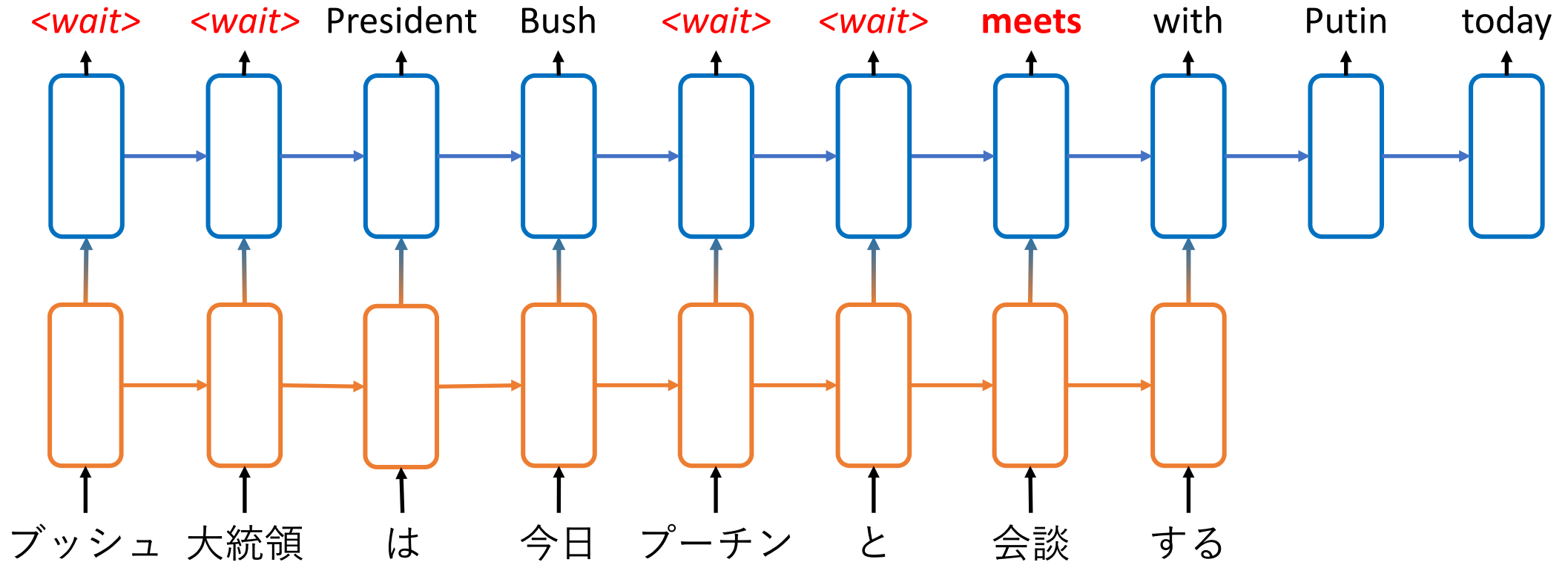
従来手法：訳出開始の固定ステップ数遅延

- Wait-k (Ma+, ACL 2019)



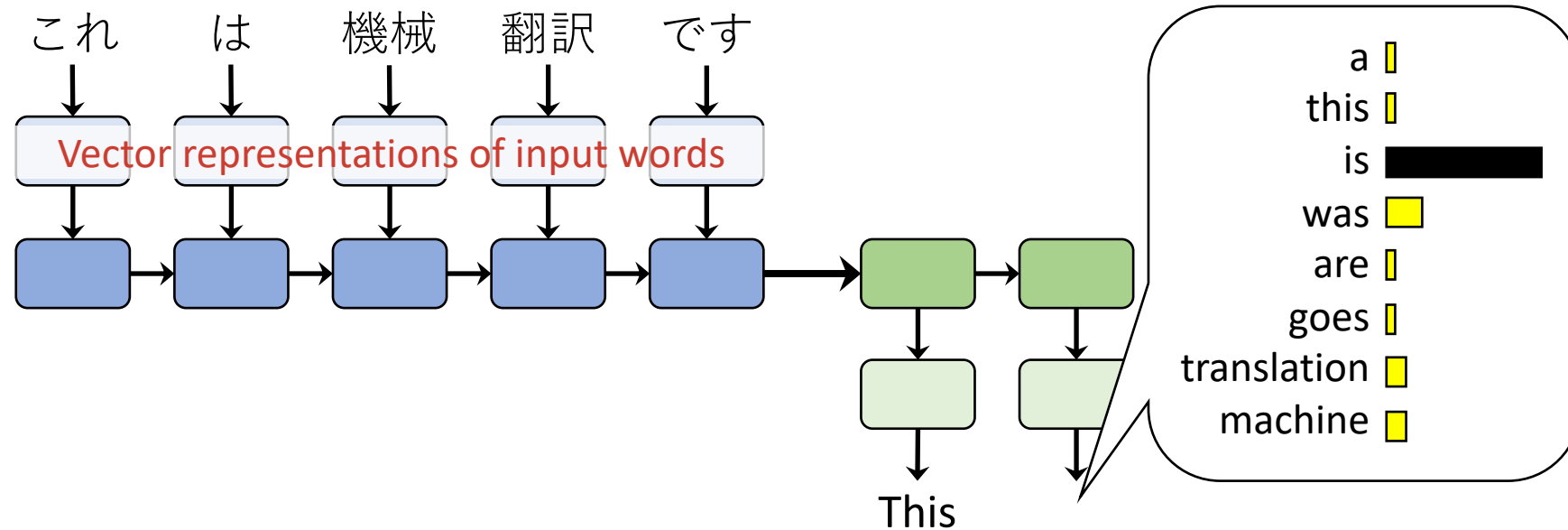
提案手法：デコーダにおける適応的遅延

- 特殊記号 `<wait>` を定義し遅延を制御



最適化

- 二種類の損失を最小化
 - 遅延（遅延発生数に比例）
 - 単語予測誤り（Softmax cross entropy）



<wait>挿入位置の問題

- <wait>の位置は自由度が高い
 - President Bush meets <wait> with
 - President Bush <wait> meets with
 - President <wait> Bush meets with
 - ...
- *Connectionist Temporal Classification* (CTC)に基づく最適化（動的計画法） [A. Graves+ ICML 2006]
 - 音声認識でよく用いられる

漸進的テキスト音声合成

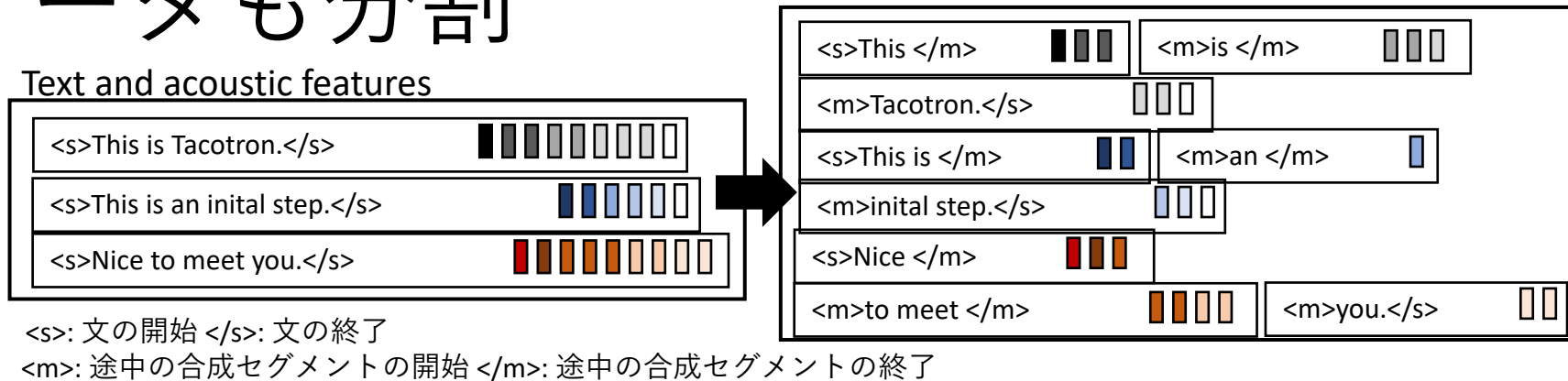
Yanagita, Sakti, and Nakamura: Neural iTTS: Toward Synthesizing Speech in Real-time with End-to-End Neural Text-to-Speech Framework, Proc. of the 10th ISCA Speech Synthesis Workshop, pp. 183-188 (2019)

テキスト音声合成と漸進的处理

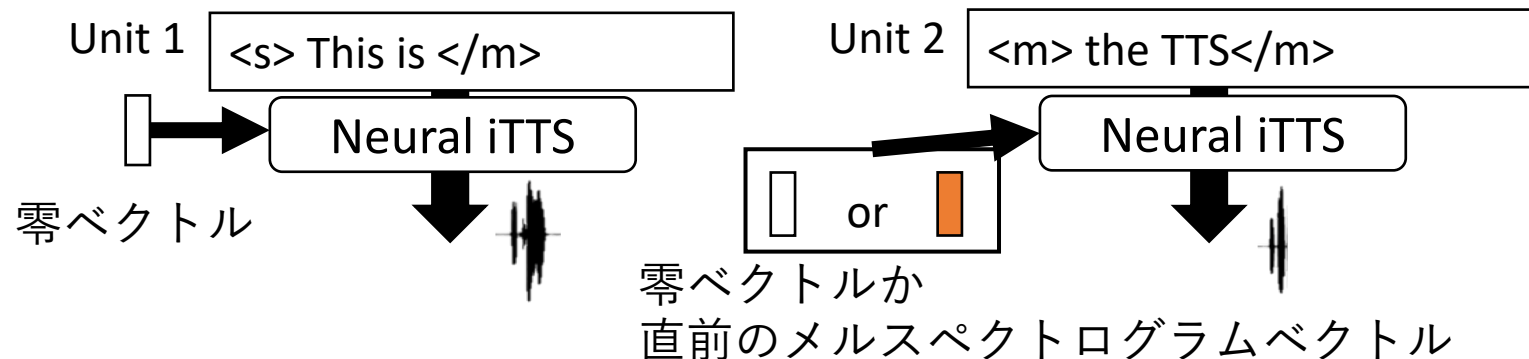
- アクセント/イントネーション決定に文脈が必要
 - 英語: 2単語程度? 日本語: 2アクセント句程度?
(提案手法における実験結果より)
- 漸進的合成でも音声を滑らかに繋ぐ
 - 分割して出力しても繋ぎ目の連続性が欲しい

提案手法: 合成過程の分割と直前結果参照

• 学習データも分割

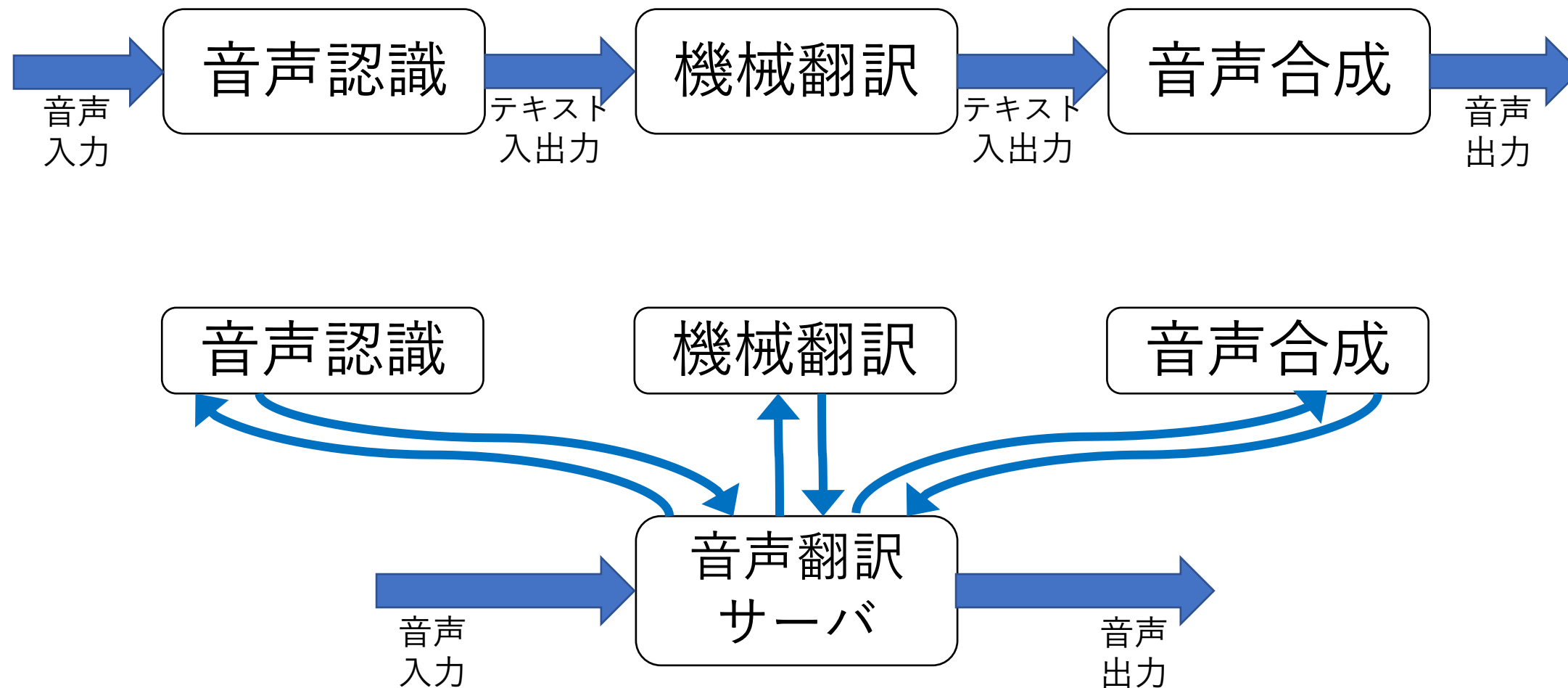


• 合成時は直前の結果を初期値として再利用



音声から音声への 同時翻訳システム

システム構成



デモ映像：同時音声翻訳



おわりに

- 同時通訳は非常に挑戦的な課題
 - 低遅延
 - 要約・簡明化
 - 予測
- 技術的に興味深い問題
 - モデルやアルゴリズム
 - 言語と知能

さらなる挑戦

- 要約・簡明化された訳出
 - 職業通訳者の通訳から学習可能か？
- 評価
 - 意味レベルの評価
 - 時間制約を考慮した実用的評価
- 通訳者の同時通訳を支援する技術