



# IWSLT Evaluation Campaign: Simultaneous Speech Translation

須藤 克仁

奈良先端科学技術大学院大学

# 同時音声翻訳

- ・漸進的な音声翻訳（≠同時通訳）



so basically the purpose of this lecture course is  
to learn basic knowledge of sequential data modeling  
that can be applied to any kind of sequential data

では 基本的に この授業の目的は  
基本知識を学ぶことです 系列データのモデル化の  
適用するのは あらゆる種類の 系列データです

# オンライン翻訳と同時通訳

(1) The relief workers (2) say (3) they  
don't have (4) enough food, water,  
shelter, and medical supplied (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』  
第141回 音声言語情報処理研究会 (オンライン)

# オフライン翻訳と同時通訳

(1) The relief workers (2) say (3) they don't have (4) enough food, water, shelter, and medical supplied (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**

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

第141回 音声言語情報処理研究会 (オンライン)

# オフライン翻訳と同時通訳

(1) The relief workers (2) say (3) they don't have (4) enough food, water, shelter, and medical supplied (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**

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

第141回 音声言語情報処理研究会 (オンライン)

# 同時翻訳の位置づけ

	同時性	原文忠実性	解釈性
オフライン翻訳	×	○	○
同時翻訳	○	○	△
同時通訳	○	△	○

# 同時翻訳の位置づけ

	同時性	原文忠実性	解釈性
オンライン翻訳	×	○	○
同時通訳	○	△	○

同時性と解釈性の両立のため  
言い換えや要約も含む高度な訳出を行う

# 同時翻訳の位置づけ

	同時性	原文忠実性	解釈性
オンライン翻訳	×	○	○
同時翻訳	○	○	△

訳出順の調整 (**順送りの訳**) のみで  
同時性を実現するため 解釈性に影響あり

# IWSLTとは？

- The International Conference on Spoken Language Translation
  - 2004年に第1回開催@ATR
  - 第17回（2020年）から呼称を conference に（略称は変更なし）
  - Evaluation Campaign（共通タスク）と Research Papers（研究発表）

# SIGSLT

- ACL, ISCA, ELRAの合同SIG
- 19名のSteering Committeeメンバー
  - Chair: Alex Waibel (CMU)
  - ACL SIG President: Marcello Federico (Amazon)
  - ISCA SIG President: Satoshi Nakamura (NAIST)
  - ELRA SIG President: Sebastian Stüker (Zoom)

# IWSLT 2022 (The 19th IWSLT)

- ACL 2022の併設会議
  - 5/26-27 ダブリン（ハイブリッド）
    - ~~3/13 Research paper due~~
    - *3/14-25 Evaluation Period*
  - Evaluation Campaign
    - **Simultaneous** / *Offline* / Low-Resource / Speech-to-Speech / **Multilingual** / Dialect / *Formality Control* / Isometric
    - 下線のものは英日(en-ja)タスクあり

# Simultaneous ST on IWSLT 2022

- <https://iwslt.org/2022/simultaneous>
- 入力の違いによる2トラック
  - Text-to-Text: ストリーミングASR入力
  - Speech-to-Text: 音声入力
- 3つの異なる目的言語
  - English-to-German (2020-)
  - English-to-Japanese (2021-)
  - English-to-Chinese (2022-)

# データセット

- 学習データ・開発データ
  - MuST-C v2.0 (Cattoni+ 2020)
    - TED Talks 英語音声 + 多言語テキスト
    - train, dev, *tst-COMMON*, *tst-HE*
    - 合計30万文、600時間程度
  - 他各種音声翻訳・機械翻訳データ
- 評価データ
  - MuST-C未公開データ

R. Cattoni et al., [MuST-C: A multilingual corpus for end-to-end speech translation](#),  
Computer Speech & Language, Vol. 66, Article 101155, March 2020.

# ベースラインシステム

- オーガナイザが構築方法を公開
  - fairseqベースの翻訳
    - en-de [speech-to-text](#)
    - en-ja [speech-to-text](#), [text-to-text](#)
    - (fairseqのバージョン不整合で難儀しました...)
  - torchaudioによるストリーミングASR
    - [Emformerモデル](#)
    - [学習レシピ](#) を公開

# 評価方法

- Dockerイメージを提出
  - オーガナイザがAWS上で実行、評価
  - 評価ツール SimulEval (Ma et al. 2020)
- 品質 (Quality) 評価
  - BLEU +一部人手評価実施予定 (en-de)
- 遅延 (Latency) 評価
  - Average Lagging

X. Ma et al., [SIMULEVAL: An Evaluation Toolkit for Simultaneous Translation](#),  
*Proc. EMNLP 2020 (Systems Demonstrations)*, November 2020.

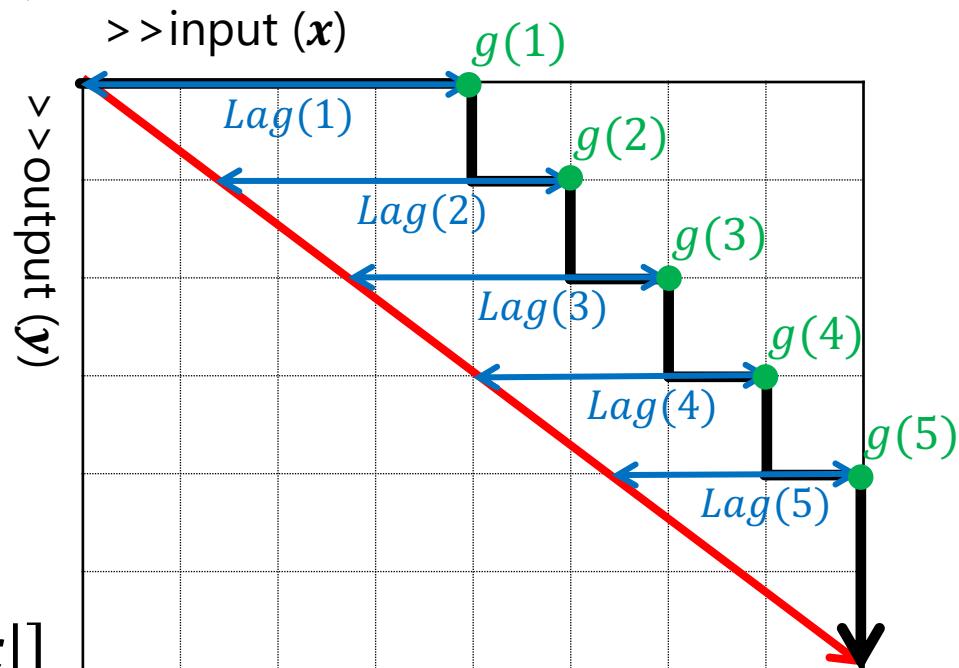
# Average Lagging

- 理想的な同時翻訳からの平均遅延

$$AL = \frac{1}{\tau} \sum_{t=1}^{\tau} Lag(t)$$

$$Lag(t) = g(t) - \frac{t-1}{\gamma}$$

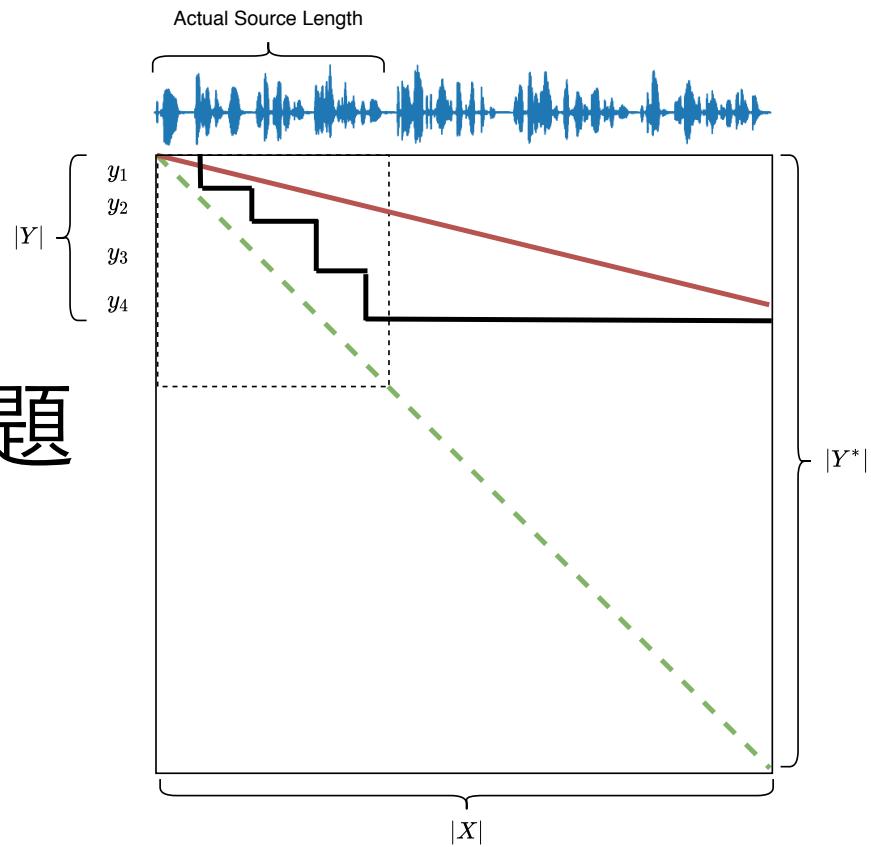
$$\gamma = \frac{|y|}{|x|} \quad \tau = \operatorname{argmin}_t [g(t) = |x|]$$



M. Ma et al., [STACL: Simultaneous Translation with Implicit Anticipation and Controllable Latency using Prefix-to-Prefix Framework](#), Proc. ACL 2019, July 2019.

# Average Lagging (SimulEval版)

- 音声入力に対応
  - 実時間による遅延
- 比較対象の修正
  - 出力が短い際の問題
  - オリジナル:  $\gamma = \frac{|y|}{|x|}$
  - SimulEval:  $\gamma = \frac{|y^*|}{|x|}$



[図は Ma et al. (2020) より - [CC BY 4.0](#)]

# 品質と遅延のトレードオフ

- 品質と遅延は反比例する
  - 意外に早く飽和？
- 複数の設定で実験比較が必要

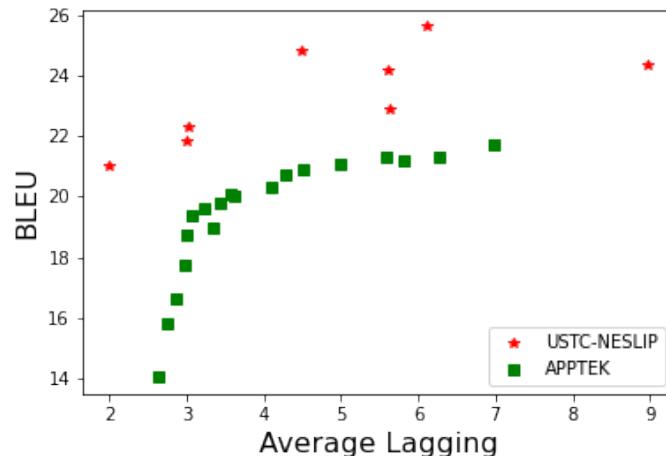


Figure 4: Latency-quality trade-off curves, measured by AL and BLEU, reported on the blind test set, for the systems submitted to the speech track with segmented input. AL is considering the computation time and measured in seconds.

図: IWSLT 2021 同時音声翻訳タスク(en-de)のAL（秒,計算時間込）とBLEUの結果  
[図は下記 Overview paper より - [CC BY 4.0](#)]

A. Anastasopoulos et al., [FINDINGS OF THE IWSLT 2021 EVALUATION CAMPAIGN](#),  
*Proc. IWSLT 2021*, August 2021.

# 遅延レジーム (Latency regimes)

- システム比較のための遅延目安
  - ドイツ語 (de) < 中国語 (zh) < 日本語 (ja)

	en-de	en-ja	en-zh
Low	≤1000 (ms)	≤2500 (ms)	≤2000 (ms)
Medium	≤2000 (ms)	≤4000 (ms)	≤3000 (ms)
High	≤4000 (ms)	≤5000 (ms)	<4000 (ms)

(en-ja, en-zh の Speech-to-speech は2022から)

# IWSLT 2021 同時翻訳参加チーム

- [UEDIN](#) (U. Edinburgh): en-de textのみ
- [VOLCTRANS](#) (ByteDance): textのみ
- [NAIST](#): en-ja textのみ
- [USTC-NELSLIP](#) (USTC+iFlytek): 全て1位
  - RNN-Tの拡張、Cross Attention Augmented Transducer (CAAT)
    - Transformerの注視機構の応用
    - end-to-end と cascadeの両方を比較
- [APPTEK](#) (AppTek): en-deのみ, cascade

# IWSLT 2021 同時翻訳結果 (en-de)

## Text-to-Text

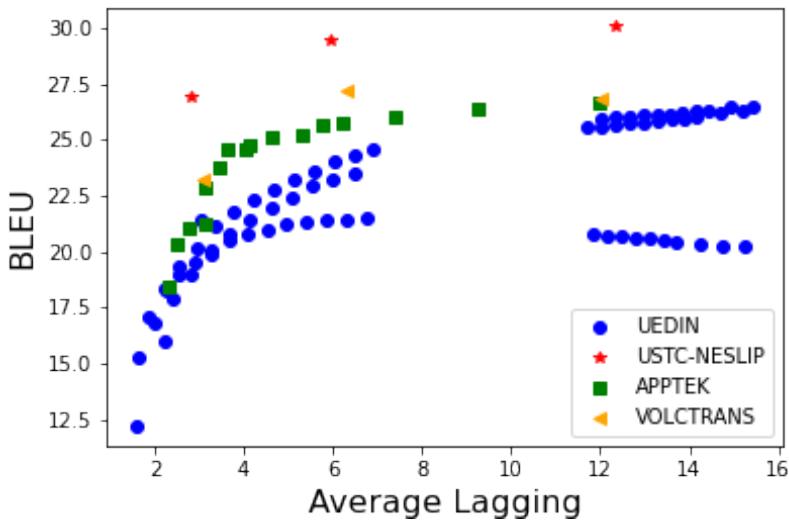


Figure 1: Latency-quality trade-off curves, measured by AL and BLEU, reported on the blind test set, for the systems submitted to the English-German text track.

## Speech-to-Text

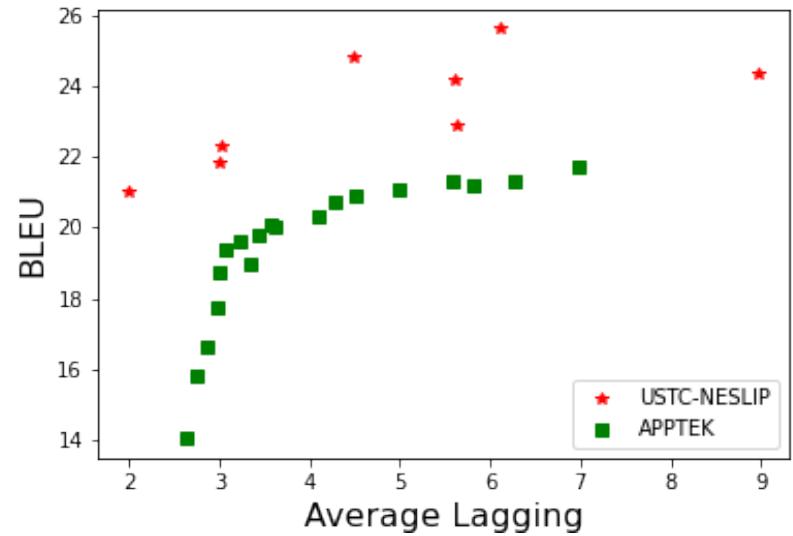


Figure 4: Latency-quality trade-off curves, measured by AL and BLEU, reported on the blind test set, for the systems submitted to the speech track with segmented input. AL is considering the computation time and measured in seconds.

[図はA. Anastasopoulos et al. (2021)より - [CC BY 4.0](#)]

# IWSLT 2021 同時翻訳結果 (en-ja)

22

## Text-to-Text

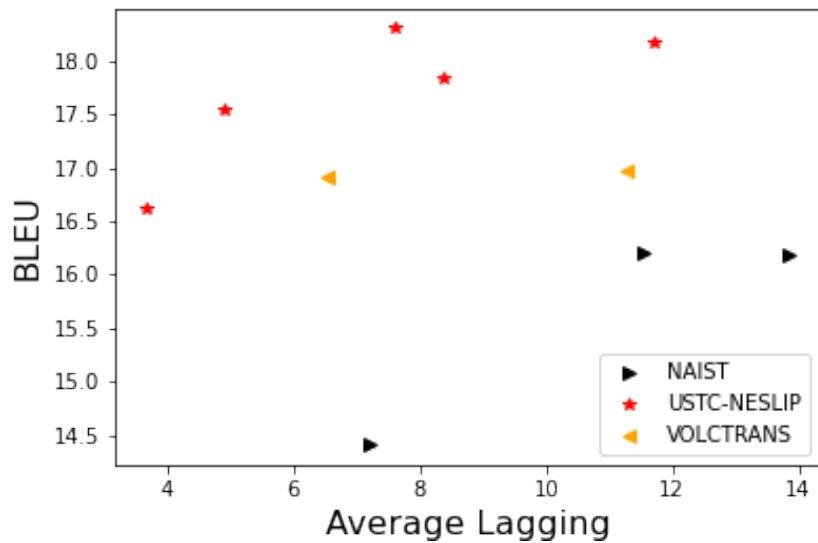


Figure 2: Latency-quality trade-off curves, measured by AL and BLEU, reported on the blind test set, for the systems submitted to the English-Japanese text track.

- データ拡張が有効
- 逆翻訳
- 知識蒸留
- 自己学習
- USTC+iFlytekの手法の効果大

[図はA. Anastasopoulos et al. (2021)より - [CC BY 4.0](#)]

# 同時翻訳の諸課題と今後の展望

- 語順の差と訳出遅延の関係
- 同時翻訳の訳質評価
- Speech-to-Speech 同時翻訳

# 語順の差と訳出遅延の関係

- 現状の評価スキームの限界
  - 学習／開発／評価データ：通常の翻訳
  - 翻訳結果：（強引な）順送りの訳
- 何が問題か？
  - 理想的な「順送りの訳」がないので遅延削減が訳質低下に直結する
  - NAIST同時通訳コーパス (Doi et al., 2021)
    - 通訳なので情報の省略や誤りもある

K. Doi et al., Large-Scale English-Japanese Simultaneous Interpretation Corpus: Construction and Analyses with Sentence-Aligned Data, Proc. IWSLT 2021, August 2021.



# 同時翻訳の訳質評価

- 参照訳が順送りでないのでBLEU等表層評価には限界がある
  - 人手評価は 2022 en-de で実施予定
  - en-ja では非公式な人手評価検討中
- 伝達内容の正しさの評価は未確立
  - 通訳研究においても同様
    - cf. 同時通訳の誤り (Doi et al., 2021)
    - ...依然として非常に挑戦的な課題

# Speech-to-Speech 同時通訳

- ・同時翻訳に限らずIWSLT全体で注目
  - ・今後の研究加速に期待
- ・漸進的ASR, MT, TTSの組み合わせでcascade型システムの構築は可能(Fukuda et al., 2021)
  - ・遅延とエラーの伝播が課題
  - ・漸進的TTSにおける発声詰まりの問題
    - ・合成音出力未了のテキストの滞積

R. Fukuda et al., [Simultaneous Speech-to-Speech Translation System with Transformer-Based Incremental ASR, MT, and TTS](#), Proc. Oriental COCOSDA 2021, November 2021.



# まとめ

- IWSLT 同時翻訳タスクの紹介
  - 実用的にも大きなチャレンジです
- タスク参加者が少ない...
  - 研究では急速にプレイヤーが増加
  - テキスト入力、音声入力とも
- 是非次回 (2023) ご参加ください
  - ACL 2023併催で提案？
  - 同時翻訳以外のタスクもあります

# Join Our Groups!

- IWSLT Evaluation Campaign
  - <https://groups.google.com/g/iwslt-evaluation-campaign>
    - 共通タスクに関する告知
- SIGSLT
  - <https://groups.google.com/g/sigslt>
    - SIGの活動に関する告知
    - ISCA SIGSLT Lectures
      - <https://iwslt.org/lectures/>