

# 同時通訳コーパスにおける単語発声時間の自動推定

Utterance Time Estimation on Simultaneous Interpretation Corpus

小野貴博<sup>1</sup>  
Takahiro Ono

松原茂樹<sup>2</sup>  
Shigeki Matsubara

名古屋大学大学院情報科学研究科<sup>1</sup>  
Graduate School of Information Science, Nagoya University

名古屋大学情報連携基盤センター<sup>2</sup>  
Information Technology Center, Nagoya University

## 1 はじめに

計算機による同時通訳の実現のために、通訳者が用いる技術を応用することは有効な方法の一つである。このような観点から、同時通訳音声の収集が行われており、百万語規模のデータベースの構築が進められている。これまで、通訳技術の分析に使用することを目的に、発話単位を最小単位とした対訳対応付けが行われたが [2]、詳細な観察をするには十分とは言えない。より細かい単位で対応付けを行うためには、細粒度で時間情報を付与する必要があるのである。

本稿では、書き起こし文への形態素単位の時間情報の付与について述べる。書き起こし文の形態素と音声信号の区間の対応を推定することにより、各形態素に開始・終了時刻を付与する。名古屋大学同時通訳データベース [1] を用いて、本手法による同時通訳データへの利用可能性を検討した。

## 2 名古屋大学同時通訳データベース

名古屋大学 CIAIR では、独話および対話の同時通訳音声データを収集してきた。収集した音声は、200ms 以上のポーズで発話単位に分割し、人手で書き起こされている。各発話単位には、ID、発話開始時間、発話終了時間、話者 ID を付与している。図 1 に、日本語独話とその英日通訳の書き起こしデータの例を示す。

## 3 形態素を単位とした時間情報付与

元となった音声信号を用いて書き起こし文に形態素単位の時間情報を付与する。

- 書き起こし文に形態素解析を行い、形態素の境界と各形態素の読みを求める。
- 音声認識システムを用いて、形態素と音声信号の区間の対応を求める。
- 音声信号の開始・終了時刻を、対応する形態素の時間情報とする。

## 4 実験と評価

本手法の有効性を評価するために、名古屋大学同時通訳データベース [1] を用いて実験を行った。日本語講演に含まれる 10 個の発話単位の書き起こし文に、本手法を用いて時間情報を付与した。形態素解析には茶筌 [3]、音声認識には Julian [4] を用いた。

各形態素について、本手法が付与した時刻と人手で付与した時刻の誤差が  $\theta$ ms 以下の場合を正解とし、正解率を計算した。 $\theta$  として、20, 40, 60, 80, 100 を設定

日本語話者発話	日英通訳者発話
0001 - 00:00:348-00:01:280 N: おはようございます<SB>	0001 - 00:01:400-00:01:959 I: Good morning.
0002 - 00:02:476-00:05:500 N: ただ今ご紹介いただきました鈴木でございます<SB>	0002 - 00:03:800-00:05:287 I: Thank you for your kind introduction,
0003 - 00:05:680-00:06:920 N: どうぞよろしくお願いします<SB>	0003 - 00:05:592-00:06:759 I: my name is Suzuki.
0004 - 00:07:728-00:13:752 N: 今日ご出席の皆様方私普段から親しくさせて...	0004 - 00:11:616-00:12:439 I: I have been
0005 - 00:14:248-00:18:336 N: (F<H>)鈴木が何を言うかということは...	0005 - 00:12:704-00:14:847 I: very good friends with the participants today.

図 1 同時通訳コーパスの書き起こし文

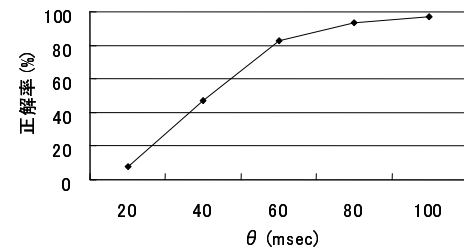


図 2 閾値  $\theta$  と正解率の関係

した。図 2 に、誤差の閾値と正解率の関係を示す。ほとんどの形態素が 100ms 以内の誤差で時間情報を付与できる。人手で時間情報を付与するときの誤差を 100msec 以下にすることは困難であると考えられ、本手法の有効性を確認した。

## 5 おわりに

本稿では、音声信号を用いて書き起こし文に形態素単位で時間情報を付与について述べた。今後は、同時通訳コーパスのフレーズ対応付与に、これらの時間情報を利用する予定である。

## 参考文献

- 遠山仁美, 松原茂樹, 河口信夫, 稲垣康善, “CIAIR 同時通訳データベースの構築と利用”, 信学技報, Vol. 103, No. 487, pp. 7-12, 電子情報通信学会, 2004.
- 高木亮, 松原茂樹, 稲垣康善, “同時通訳コーパスの対訳アライメント手法とその評価”, 情報処理学会第 64 回全国大会講演論文集, 2002.
- 松本裕治, 北内啓, 山下達雄, 平野善隆, 松田寛, 高岡一馬, 浅原正幸, “形態素解析システム『茶筌』 version 2.3.3 使用説明書”, 奈良先端科学技術大学院大学松本研究室, 2003.
- 大語彙連続音声認識システム Julius, <http://julius.sourceforge.jp/>