Demonstrating the Effectiveness of Speech Acquisition and Text Conversion Accuracy in Multiple Parallel Group Work in a Large Classroom

大教室での複数並列グループワークにおける，音声取得とテキスト変換精度の有効性の実証

Demonstrating the Effectiveness of Speech Acquisition and Text Conversion Accuracy in Multiple Parallel Group Work in a Large Classroom

小笠原　豊，大橋　一広，
株式会社イトーキ　DX 推進本部　デジタルソリューション企画統括部

概要

大教室で複数チームが同時に並列にて行う対面式のグループワーク，および対面参加者と遠隔参加者で構成されるハイフレックス型のグループワークにおいて，グループワークの記録および活動データ利活用の観点から，ウェブ会議システムを用いてグループワーク各々のチームの音声（動画）を記録する場合，空間の背景騒音の影響で音声のテキスト変換精度が低くなることに課題がある．

本稿では，対象とする１つの対面式グループワークの音声の取得にあたり，高性能ビデオサウンドバー（以下，ビデオサウンドバーと呼ぶ）[1]を用いる方法と，学生のノートPC(BYOD)内蔵マイクによる方法とで比較し，実際の授業が行われる大教室の環境において，音声のテキスト変換精度を検証する実験を行った．実験の結果，音声認識の精度に差がみられ，ビデオサウンドバーを用いることが，音声のテキスト化精度を向上させることが分かった．

キーワード： グループワーク，音声データ活用，音声のテキスト変換

1. はじめに

株式会社イトーキは，武蔵野大学中長期計画の中で，”チャレンジ３「AI 世界を先導するMUSIC”というテーマを受けて立ち上がった MUSIC 計画推進小委員会の下部組織としてのタスクフォース活動 SI 響室の構築活動に参画し，学修環境の高度 IT 化および，学修環境から取得できる，学生の活動データの利活用の検討・実証を行っている．その対象として考えるデータの一つとして，グループワークにおける音声データがある．これからの社会に求められる課題解決の資質，能力を涵養するための学びの方法として，グループワークの重要性は増してきており，グループワークの音声データを記録，利活用し，学修支援，および

教員支援につなげることが重要であると考える．そのためには，グループワークにおいてウェブ会議システムを用い，学生の会話を録音し，音声をテキスト変換したデータを取得する必要があるが，大教室のように背景騒音が多い環境では，記録した音声のテキスト変換精度が低いことに課題がある．こうした課題を解決する方法として，グループワーク毎に，高性能のビデオサウンドバーを用いることが考えられ，その音声取得と変換精度について実験を行った．

2. 方法

2.1 検証対象授業について

2022 年度のサービスデザイン[a]の授業内において，学生には「Happy なゴミ箱」をつくるという課題が与えられた．学生はその課題を解くために，１つの広い教室（武蔵野大学 有明校舎４号館 303 教室）において，1 チーム 4–5 名で，十数のグループに分かれ，最終アウトプットを行うまでの一連の作業（対面でのディスカッション，アイデア出し，教員からの指導と再考の繰り返しによるプロトタイピング，アイデア収束，発表資料作成，発表練習から最終発表まで）を行なった．

図１　サービスデザイン[a]授業風景

Figure 1 Scene from the Service Design [a] class

2.2 実験方法

上記の対象授業において，1 つのグループワークチームを対象とし，ノート PC で音声を取得する方法と，ビデオサウンドバーを使用して音声を取得する方法，この 2 つの方法で同じグループワークの会話を録音（ノート PC での録音は学生によるもの）した．ビデオサウンドバーは，録音の集音範囲設定することが可能で，対象グループワークの音声をカバーする距離，本体から 2.5m 以内の音を集音する設定とした．録音のプラットフォームと

して Microsoft TEAMS[2]のビデオ会議機能を使用した.

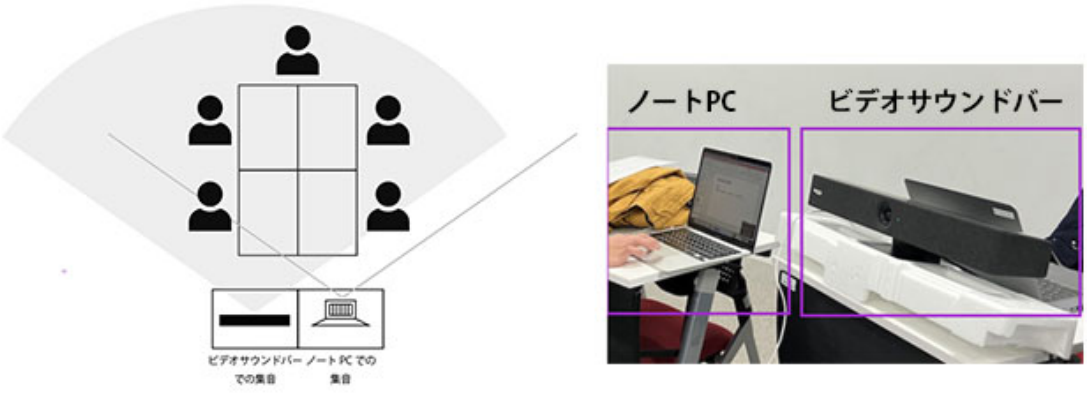

図 2　対象のグループワークとマイクの配置イメージ

Figure 2 Image of target group work and microphone placement

また，取得した音声を，Microsoft Stream[3]のトランスクリプト生成機能を用いてテキスト化し，認識率の比較を行った．音源は，2022 年 12 月 15 日に収録した学生の発表練習時の音源（A）と，2022 年 12 月 22 日に収録した教員とのディスカッション時の音（B)，の異なるシチュエーションにおける，それぞれ 5 分程度の会話である.

2.3　データセット

テキストの変換精度を比較するための基準となるデータは，上記の音源(A)音源（B）の録音から人力で文字起こしを行なったものを使用した．また，音源（A）および，音源（B）については，ビデオサウンドバーによる録音，ノート PC による録音のそれぞれについて，Microsoft Stream のトランスクリプト機能を用いてテキスト化し実験データとして用いた.

2.4　評価方法

取得した音声の分析には，機械翻訳の評価方法として広く使用されている BLEU スコアを用いた．BLEU の計算結果は 0～1 の間の少数として出力され，その結果の数値を 100 倍しスコアとして表現する．この数値が高くなるにつれ品質が高くなり，一般的に，スコアが 40 以上であれば高品質であるとされる.

表 1　BLEU スコアの解釈[4]

Table 1 Interpretation of BLEU Scores [4]

BLEU スコア	解釈
< 10	ほとんど役に立たない
10〜19	主旨を理解するのが困難である
20〜29	主旨は明白であるが、文法上の重大なエラーがある
30〜40	理解できる、適度な品質の翻訳
40〜50	高品質な翻訳
50〜60	非常に高品質で、適切かつ流暢な翻訳
> 60	人が翻訳した場合よりも高品質であることが多い

3. 結果

以下が，ビデオサウンドバー，およびノート PC を使用して音声を取得，テキスト化した場合の BLEU スコアの比較結果である．ビデオサウンドバーを使用して音声を取得した場合の BLEU スコアは，「理解できる，適度な品質の翻訳」とされるレンジである 35〜40 に位置する一方で，ノート PC での取得の場合，16〜26 と「主旨を理解するのが困難である」「主旨は明白であるが, 文法上の重大なエラーがある」といった 1〜2 段下のレンジに属し，ビデオサウンドバーによって取得した音声をもとにしたテキスト変換の方が，精度が高いことが確認された．

表 2　BLEU スコア計算結果

Table 2 BLEU Score Calculation Results

	BLEU スコア(音源 A)	BLEU スコア(音源 B)
ビデオサウンドバー	40	35
ノート PC	26	16

4. 考察

ビデオサウンドバーによる録音については，その機能（集音する空間の範囲を，マイクからの鉛直距離(実験では 2.5m までとした)と，水平方向の角度（110 度まで）の組合せで決め，範囲外の音声を収録しないように制御でき，またサウンドバー内蔵のカメラによる人認識機能と連動するマイクが，鋭い指向性で発話者の音声を収録できる）によって，他のグループワークの発話に起因する背景騒音をある程度抑制しながら，より明瞭に発話者の音声収録を行うことができたと考えられる．

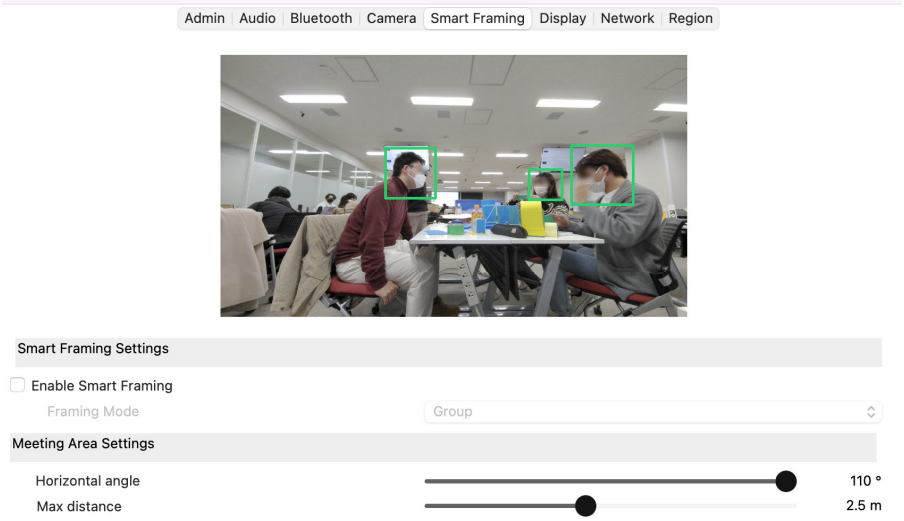

図 3　ビデオサウンドバーの機能画面（顔認識の緑枠はイメージ）

Figure 3 Video Soundbar function screen

(The green frame with face recognition is an image)

これに対してノート PC での録音音源は，TEAMS のノイズキャンセル機能によって，背景騒音が低い状態で収録されており，一見したところ聞きやすく感じるが，反面，ノイズキャンセルの影響か，随所で収録した発話に崩れが起きる事によって，正確な言葉を捉えづらくなっている面が見受けられる．そしてその事が，同時に音声のテキスト変換の精度に影響を与え，変換精度の低下を招いていると考えられる．

5.　おわりに

近年の音声のテキスト変換の精度向上により，文字起こしデータの活用における有用性を増してきているが，収録する音そのものが良くなければテキストの変換精度も損なわれる．本実験の結果より，対面でのグループワークにおいて音声を取得する場合，高性能なビデオサウンドバーを用いることで，聴覚的な聞きやすさを担保しつつ，音声のテキスト変換の正確性をも確保できることが分かった．

図 4　ビデオサウンドバーを使ったハイフレックスグループワーク

Figure 4 high-flex group work with Video Sound bar

こうして，良好な変換精度のテキストデータを残せる事によって，発話内容，発言量，重要語などを用いた，テキストマイニングによるグループワークの振り返り支援への活用，例えば，前回不参加だったメンバーが事前に活動内容を確認することや，あるいはメンバー全員で，映像とテキストを同時に参照しながら，グループワークの開始後，数分で前回を振り返る事が可能となっていく．

学修活動の記録として，音声のより正確な取得から，テキスト化を通じたフィードバック，学修の振り返り支援へと，データの取得・活用の道筋を引き続き検討していきたい．

図 5　グループワーク振り返りアプリケーション

（データサイエンス学部中西研究室と共同開発中）

Figure 5 Group Work Reflection Application

(Under development in collaboration with Nakanishi Laboratory, Faculty of Data Science)

謝辞

本実験の機会をいただきました MUSIC センター長，上林憲行教授，田丸恵理子先生，大崎理乃先生に感謝申し上げます.

また末尾に記載いたしました，グループワーク振り返りアプリケーションの共同開発にご尽力いただいている，データサイエンス学科長　中西崇文准教授，岡田龍太郎助教，および，中西研究室の学生の皆様に感謝いたします.

参考文献

[1]今回の実験は，高機能ビデオサウンドバーとして，機種を選考し YAMAHA 製 CS-800 を採用し実証した.

YAMAHA CS800　https://sound-solution.yamaha.com/products/uc/cs-800/index

[2]Microsoft TEAMS https://www.microsoft.com/ja-jp/microsoft-teams/group-chat-software

[3]Microsoft Stream　https://www.microsoft.com/ja-jp/microsoft-365/microsoft-stream

[4]BLEU スコアの解釈　https://cloud.google.com/translate/automl/docs/evaluate?hl=ja#bleu