

# 講師映像と解説音声の品質向上のための 講義映像変換システムの実現

松浦 辰雄<sup>†</sup> 大園 忠親<sup>†</sup> 新谷 虎松<sup>†</sup>

名古屋工業大学大学院情報工学専攻<sup>†</sup>

## 1 はじめに

反転授業の導入や講義映像配信サービスの増加により講義映像の需要が高まっている。しかし、教師による講義映像の撮影や編集のコストは高い。また、講義映像のオンライン配信により、多くの生徒が視聴でき、生徒の要望や好みは多様化する。本研究では、教師は半自動的に模範的な講義映像に変換でき、生徒は個人の好みに合わせて講義映像を変換できる講義映像変換システムを実現した。本稿では、本システムの実行例を示し、講義映像の品質向上の手法について述べ、解説音声の品質について評価し考察する。

## 2 講師映像と解説音声の品質向上

講師映像と解説音声の品質向上について述べる。反転授業において、生徒は自宅で講義映像を用いて学習を行う。また、JMOOCやKhan Academyなどの講義映像配信サービスも登場し、講義映像の需要が高まっている。しかし、講義映像の撮影や編集コストは高く、講義映像の作成支援は必要である。先行研究では、教師の講義映像の作成の支援を行なった [1]。また、講義映像のオンライン配信により、多くの生徒が視聴できるようになり、生徒の要望や好みは多様化する。そこで、本研究では、教師側では、講義映像の編集を支援し、生徒側では、個人の好みに合わせた講義映像の変換を実現する。また、多くの講義映像では、PCにスライドや作業画面を映し、講義を解説するため、解説音声が重要となる。本研究では、教師の解説音声の音量、発話速度、無声休止および有声休止を評価し、解説音声の品質向上を行う。さらに、解説音声以外の講師映像は、生徒の講義映像の視聴履歴を教師にフィードバックし、品質向上を図る。

### Realizing a Lecture Video Converter to Improve Appearance and Speech

<sup>†</sup>Tatsuo MATSUURA, <sup>†</sup>Tadachika OZONO and <sup>†</sup>Toramatsu SHINTANI

<sup>†</sup>Dept. of Computer Science, Graduate School of Engineering, Nagoya Institute of Technology.



図 1: 教師の講義映像の編集

## 3 講義映像変換システム

本システムの実行例を示しながら、本システムの特長3点を述べる。まず、教師の編集コストの削減が期待できる点である。図1(A)では、教師が解説音声の品質を元に半自動的に模範的な講義映像に変換できる。次に、生徒の視聴履歴が可視化され、教師は、視聴履歴を見直ししながら、講師映像の改善ができる点である。図1(B)では、解説音声の品質および生徒の視聴履歴を参考にしながら、講義映像の編集ができる。最後に、生徒自身の好みに合わせて、講義映像を編集することができる点である。図2(A)では、解説音声の品質を元に交換操作ができる。例えば、素早く講義内容の概要を把握したい生徒は、発話速度を高速化し講義映像を効率良く視聴できる。

解説音声の品質の評価および変換について述べる。音量  $V$  は、講義映像中の音量  $v$  に対して、周波数  $H$  を考慮し、dB 単位で評価する (式 (1))。

$$V_i = v_i + 1.7 \times 10^{-2} H_i - 1.4 \times 10 \quad (1)$$

$V$  が 45dB を満たしていない場合、音量正規化を行う。発話速度は、1分当たりのモーラ数  $M$  を  $t$  の音声区間の時間および  $p$  の音素数から評価する (式 (2))。

$$M_i = 60 \times \frac{p_i}{t_i} \quad (2)$$

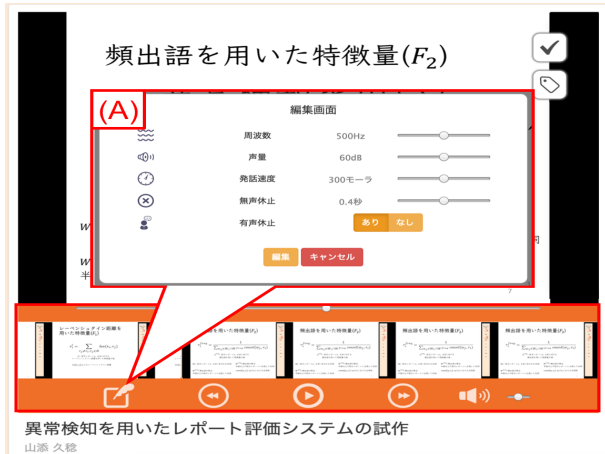


図 2: 生徒の講義映像の視聴

1分あたり300mora以上の場合、PSOLAを用いて、周波数を変えずに音声の伸縮を行う。無声休止は、横井ら[2]より、接続詞、体言止めおよび名詞+助詞の後に400ms取られているか評価し、400ms未満の場合は、無音を挿入する。有声休止は、後藤ら[3]より、基本周波数およびスペクトル包絡の変化量から見本の有声休止を評価する。ここでは、見本 $a$ との類似度が高い音声 $i$ を、有声休止とした。類似度 $S$ として式(3)のように余弦を用いた。

$$S_i = \frac{M_a \cdot M_i}{||M_a|| ||M_i||} \quad (3)$$

ここで、 $M_a$  および  $M_i$  は、それぞれ見本  $a$  および評価対象音声  $i$  のメル周波数ケプストラム係数ベクトルである。

## 4 評価実験

### 4.1 模範的な解説音声の品質の検出精度

模範的な解説音声の品質の検出精度を評価する。被験者5人がPCとマイクを用いて、過去に学会で発表した内容を5分間収録した。被験者には自身以外の講義映像に対して、5秒ごとに、音量は、十分な声の大きさか、発話速度は、十分聞き取れる速さか、無声休止は、十分に無声の間があるかおよび有声休止の有無を判定させた。そして、本システムの検出結果と被験者全員分の判定データを比較した(表1)。講義映像はマイクなどの収録により、音量が確保されるため、本実験においても音量不足と判定された箇所は少なく、音量を考慮する優先度は低いと考える。発話速度および無声休止は、高い再現率を得ており、解説音声の早口の箇所を網羅的に検出でき、正しく修正できると期待できる。有声休止は、高い適合率を示し、正しく有声休止を削除でき、解説音声の品質の向上が期待できる。

表 1: 評価実験結果

	適合率	再現率	F 値	正確度
音量	0.0656	0.667	0.119	0.940
発話速度	0.603	0.761	0.673	0.650
無声休止	0.506	0.837	0.631	0.520
有声休止	0.701	0.641	0.670	0.576

表 2: 解説音声の好みの調査

	A	B	C	D	E
音量 (dB)	58	60	63	61	57
発話速度 (mora)	367	332	333	424	250
無声休止 (ms)	890	210	940	30	700
有声休止 (ms)	56	145	95	25	25

### 4.2 解説音声の好みの調査

生徒の解説音声の好みの調査した。本実験結果における各被験者における最高F値をとる閾値を表2に示す。特に、被験者Dは、発話速度のモラ値が高く、無声休止は短いため解説音声の高速再生を好んでいると考えられる。対して、被験者Eは、解説音声の低速再生を好んでいると考えられ、解説音声の再生速度に好みが大きく分かれると考えられる。また、有声休止の長さにおいて、生徒によって判定する基準が異なる。被験者Bは、他の被験者に比べて有声休止が気になりにくいと考えられる。以上より、生徒の解説音声の好みの多様性が確認でき、本システムによる生徒個人が講義映像を変換するインターフェースが有用であることが示唆された。

## 5 おわりに

本研究では、講師映像と解説音声の品質向上のための講義映像変換システムを実現した。本システムでは、教師側は模範的な講義映像、生徒側は個人の好みに合わせた講義映像に変換できる。

### 謝辞

本研究の一部は JSPS 科研費 JP15K00422, JP16K00420 の助成を受けたものです。

### 参考文献

- [1] 松浦辰雄, 大冨忠親, 新谷虎松, “反転授業における音声明瞭性を考慮した講義映像作成支援システムの開発”, 第16回情報科学技術フォーラム, vol.16, no.4, pp.81-84(2017)(FIT論文賞受賞).
- [2] 横井聖宏, 馬場康輔, 須藤秀紹, 山路奈保子, “発話中の「間」がプレゼンテーションに対する聴衆の支持に与える影響 書評ゲーム『ビブリオバトル』の発音音声録音データ分析による考察”, 日本感性工学会論文誌, vol.15, no.3, pp.363-368(2016).
- [3] 後藤真孝, 伊藤克互, 速水悟, “自然発話中の有声休止箇所のリアルタイム検出システム”, 電子情報通信学会論文誌, vol.83, no.11, pp.2330-2340(2000).