

顔面の動きをフィードバックとして用いる発話訓練システムの検討*

○北村達也 (甲南大) 川村直子 (姫獨大) 能田由紀子 (ATR-P) 吐師道子 (県広大)

1 はじめに

大学生・院生の約3割が日常的に発話のしにくさを自覚し、そのほとんどは発話を改善したいと考えている [1]. そこで、我々は簡便、安価、かつ効果的な発話訓練法を検討している. その1つである顎運動制約法は、前歯で割り箸などを噛んだ状態で練習する方法であり、ボイストレーニングの現場でも利用されている. この訓練法の効果の検証 [2][3], ボイストレーナーが実際に訓練する様子の分析から、(1) 前歯で割り箸を噛みながらも下顎以外の部分を十分に動かすこと、(2) 声を十分に出すこと、(3) トレーナーが適宜フィードバックすることが重要との認識に至った. 本研究では、これらの要因を考慮した発話訓練システムを開発し、その効果を評価する.

2 発話訓練システム

本研究では顔面の動きと音声の大きさをリアルタイムにフィードバックするシステムを開発した. システムのスクリーンショットを図1に示す. 画面上部には、訓練用の文、顔面の動きと声の大きさを表すレベルメーターなどが表示される.

顔面の動きとして、カメラ画像から算出したオプティカルフロー (OF) における顔領域の総移動量を用いた. 同時に、OF そのものも表示することにより、訓練者は自分の顔のどの部分が動いているか/動いていないかを把握することができる. 音声の大きさは、マイクから取得した音声のレベルを表示した.

訓練中、顔面の動きと声の大きさを維持させるため、レベルメーターの値が閾値を一定時間下回ると、口の動きもしくは声を大きくするよう促すメッセージが表示される. これはボイストレーナーによる声かけに対応するものである.

3 実験の方法

3.1 実験参加者

日本語を母語とする 20 歳から 22 歳の男性 12 名、女性 2 名、計 14 名が実験に参加した.

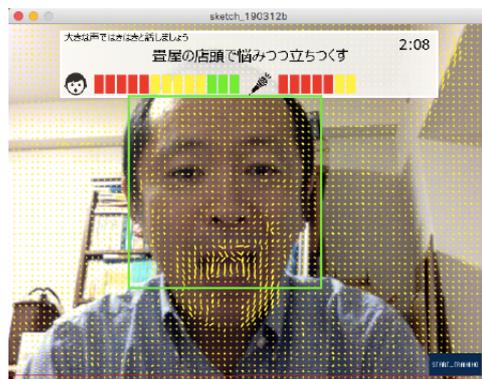


Fig. 1 Screenshot of the proposed speech training system. Green box and yellow lines indicate the facial area and optical flow, respectively.

3.2 発話資料

文献 [4] を参考にして、カ行、ガ行、サ行、ザ行、タ行、ダ行、ラ行の 5 音 (例えばカ行ならカキクケコ) を含む文を各行につき 2 文、計 14 文用意した. カ行の 1 文は「個々のキツツキの比較結果が気にかかる」である. 2 文のうちの 1 文、計 7 文を訓練に用い、訓練前後の音声の収録には 14 文を用いた.

3.3 実験手続き

実験は防音室にて行った. 実験に先立ち、実験参加者に実験の説明を行い、同意書に署名を得た. 次に、実験参加者に発話訓練の意義を説明し、意欲を持って実験に参加するよう依頼した. 声量や話速は、高校の教室で朗読することをイメージするよう指示した.

音声収録および発話訓練は立位にて行った. 訓練時は、PC のディスプレイ (EIZO EV2450) を実験参加者の顔の正面にくるよう配置し、正面を向いた状態で練習できるようにした.

実験では、まず実験参加者の訓練前の音声を録音し、Visual analog scale (VAS) にてその発話がどの程度うまくできたかを自己評価させた. 続いて、前歯で割り箸を噛んだ状態で、上記の発話訓練システムを用いて 3 分間練習した. その後、訓練後の音声を収録し、再び VAS を計測

*Speech training system using facial movement as feedback by KITAMURA, Tatsuya (Konan Univ.), KAWAMURA, Naoko (Himeji Dokkyo Univ.), NOTA, Yukiko (ATR-Promotions), and HASHI, Michiko (Pref. Univ. Hiroshima)

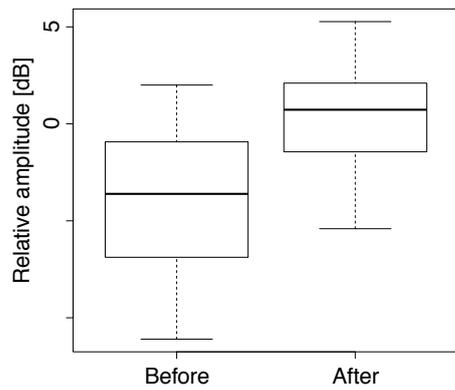


Fig. 2 Amplitude of speech recorded before and after training.

した。音声はコンデンサマイクロフォン (Sony ECM-77B) とレコーダ (Marantz PMD671) を用いて標本化周波数 16 kHz, 量子化 16 bit にて収録した。

以上の実験手続きは, 甲南大学の倫理委員会による承認を受けている。

4 結果と考察

訓練前後に録音した音声の振幅と基本周波数の変化幅をそれぞれ図 2, 図 3 に示す。これらの図は各実験参加者の 14 文の平均値の分布を示している。この結果から訓練により振幅と基本周波数の変化幅がともに上昇する傾向にあることがわかる。訓練前後の中央値の比較では, 振幅が 4.3 dB, 基本周波数の変化幅が 1.19 semitone 上昇した。図は示さないが, 基本周波数の平均値も訓練後に上昇する傾向にあった。訓練時に音声のレベルをフィードバックしたことにより, 振幅や高さの抑揚が大きくなる効果があったと考えられる。

訓練前後に計測した VAS を図 2 に示す。ほぼ全ての実験参加者が, 訓練によってうまく読めるようになったという自覚を持った。実験後には「ハキハキ言えるようになった」, 「サ行, タ行が良くなった」, 「(口の) 横の筋肉が動きやすくなった」などの肯定的なコメントが多く聞かれた。顔面の動きをフィードバックすることにより, わずか 3 分間の練習でも口の動きを改善する効果があったと考えられる。

5 おわりに

本研究では顔面の動きを利用した発話訓練システムを提案し, その効果を調査した。今後, 滑

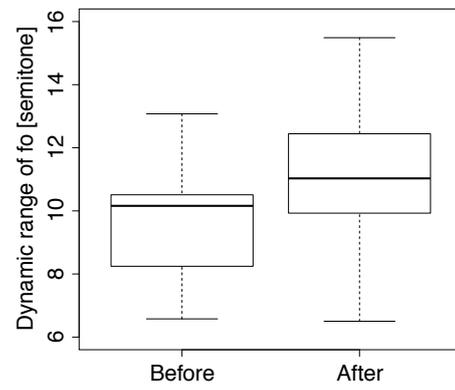


Fig. 3 Dynamic range of the pitch frequency of speech recorded before and after training.

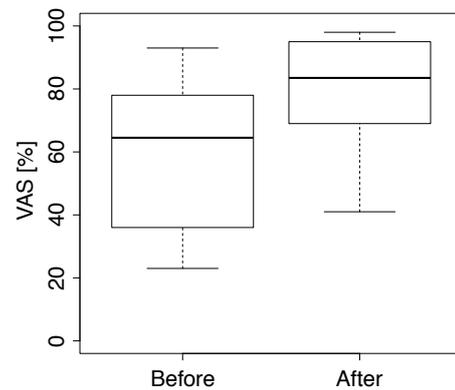


Fig. 4 Visual analog scale measured before and after training.

舌の向上に関する分析, 継続的な訓練による効果の調査を行う必要がある。

提案法はカメラとマイクロフォンがあれば実装できるので, PC のみならずタブレットやスマートフォン上に実装し, いつでも簡単に練習する環境を作ることができる。また, 発話訓練システム上にお手本の顔画像を示すことにより訓練の効果が増す可能性がある。さらに, Castillo ら [5] のようにロボットを活用して訓練者の意欲を引き出すことも考えられる。

謝辞 本研究は, 科研費 (No. 16K13226), 2019 年度ひょうご科学技術協会, 2019 年度カワイサウンド技術・音楽振興財団の支援により行われた。

参考文献

- [1] 北村ら, 音響誌, 75(3), 118-124 (2019).
- [2] 三谷ら, 音講論, 389-390 (2017 年 9 月).
- [3] 三谷ら, 音講論, 1387-1388 (2018 年 3 月).
- [4] 塩原, 声を出して読む日本語の本, 創拓社 (1987).
- [5] Castillo *et al.*, *J. Health. Eng.*, 7075290 (2018).