

顔の動きをフィードバックする発話訓練支援システムの検討*

☆設楽 郁巳, △関 和広, 北村 達也 (甲南大)

1 はじめに

医学的には健常に分類される大学生, 大学院生の約3割が日常的に発話のしにくさを感じ, そのほとんどが発話の改善を希望しているという報告がある [1]. また, 鴨居と沼田 [2] は発音, 発声, 滑舌などを学ぶことによって, 不安や心配, 自信がない等の人前で話すことに対するネガティブな意識が軽減されると報告している. これらの背景から, 我々は簡便かつ効果的な発話訓練法について検討している.

アナウンス学校等では, 割り箸などの棒状のものを横にくわえながら発話を行う訓練法が用いられている. 開口を制約して発話訓練することにより開口度の増大や調音空間の拡大などの効果をもたらすと考えられるが, その効果や理由は明確に示されていない [3].

Tang ら [4] は Clear speech 発話時の顔ランドマークの動きは Plain speech 発話時より大きいことを示している. そこで, 北村ら [5] は発話時の顔のオプティカルフローの大きさをフィードバックする発話訓練支援システムを提案した. このシステムはユーザの顔の動きの大きさをカメラ映像から算出し, リアルタイムに表示する. しかし, このシステムは計算量が大きいため動作が高速とは言えず, 移植性も低い.

本研究では, これらの問題点を克服した発話訓練支援システムの開発を行い, 開口を制約する訓練法と組み合わせて予備的な評価実験を行った.

2 発話訓練支援システム

本研究では発話中の顔ランドマークの動きの大きさをリアルタイムに表示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. 画面中央には Web カメラで撮影した利用者の顔が表示され, その上にランドマークの位置が示される. 画面右上部にはランドマークの動きの大きさを表すメーターが表示される. 画面上部にはタイマーと練習文が配置され, 練習文はあらかじめ設定された時間ごとに切り替わる.

顔の動きの計測には, 機械学習ライブラリ Me-

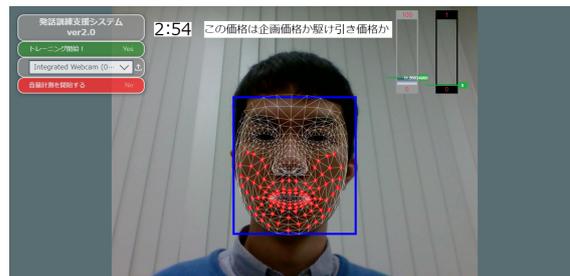


Fig. 1 Screenshot of the proposed vocal exercise support system.

diaPipe の Face Mesh を使用した. Face Mesh は画像や動画上の顔を検出し 468 点のランドマークを推定することができる. 本研究では発話運動に関係すると考えられる口唇, 頬, 顎の 106 点を抽出し表示している. 図1において利用者の顔に表示されている赤い点は, Face Mesh が推定したランドマークである. Face Mesh は顔以外の部分に関しては処理を行わないため, 動作が高速である.

上記の 106 点のランドマークの現在の座標と 1 フレーム前の座標より, 各点の移動量を計測した. そして, 全ランドマークの移動量の総和を画面左上のメーターとして表示させている. 利用者はこのメーターが高い値を示し続けるよう意識しながら発話訓練を行う. 発話と関係のない頭部の平行移動と回転移動に関しては, 鼻根点 (ナジオン) を中心にしてユークリッド変換により補正をした.

本システムは Javascript および HTML により記述され, 仮想サーバー上に配置した. 利用者はインターネットに接続し Web ブラウザが使える環境さえあれば, 本システムを利用できる.

3 評価実験の方法

提案システムおよび開口制約に基づく発話訓練法の効果について予備的な実験を行った.

3.1 実験参加者

20 歳から 22 歳の大学生計 16 名 (男性 14 名, 女性 2 名) が実験に参加した. 彼らは日本語を母

*Vocal exercise support system using feedback of facial movement. by SHITARA, Ikumi, SEKI Kazuhiro and KITAMURA, Tatsuya (Konan Univ.)

Table 1 Combination of vocal exercises performed by each group.

Groups	System	Chopstick
A	—	—
B	—	○
C	○	—
D	○	○

語とし、過去に医師に発話や聴覚について指摘されたことがない。

3.2 発話資料

カ行、ガ行、サ行、ザ行、タ行、ダ行の5音をすべて含む文を各行につき2文、合計12文用意した。各行の2文のうち1文、計6文を訓練の際に読み上げる文とし、訓練前後の読み上げにはすべての文を使用した。

3.3 手続き

実験に先立ち、実験参加者に実験説明を口頭と文書で行い署名による同意を得た。

実験参加者は4名ずつ4群に分けた。発話訓練の組み合わせを表1に示す。発話訓練支援システムを使用しないA群、B群には、PCの画面中央に文を表示し、一定間隔ごとに切り替わるプログラムを使用させた。B、D群では、発話訓練の前にくわえ方の例を示す画像がPCの画面に表示される。

実験はリモート形式で実施した。実験参加者に実験で使用するプログラムが配置されているサーバーのURLを伝え、自宅からPCで接続するよう指示した。実験参加者にICレコーダー(型番:ICD-UX523F/ICD-UX533F)とマイクロフォン(型番:ECM-SP-10)を渡し、実験中の音声を録音するよう指示した。

実験ではまず、訓練前に文の読み上げをさせた。そしてGoogle formsによってその発話がどの程度うまくできたかを10段階で自己評価させた。続いて、3分間の発話訓練を実施した。最後に、訓練後の文の読み上げと、自己評価を行わせた。この手続きを1日1回、3日間連続で実施した。1回に要する時間は訓練前後の読み上げを合わせ約10分である。

3.4 分析方法

本稿では、実験参加者が送信したGoogle formsの自己評価点を元に、主観的な訓練効果について

Table 2 The average of the difference of the self rated score between before and after three-day vocal exercises.

Groups	A	B	C	D
Δ Score	1.5	1.3	3.0	3.0

て分析を行った。

4 評価実験の結果

各群の1日目の訓練前における自己評価点と、3日目の訓練後の自己評価点との差分を平均した値を表2に示す。システムを使用しなかったA群、B群に比べ、システムを使用したC群、D群の自己評価の増加が大きい傾向が見られた。システムの利用によって、少なくとも主観的な効果を感じられる可能性がある。一方で、割り箸を利用した際の自己評価には大きな変化は見られなかった。

5 おわりに

本研究では、発話時の顔の動きの大きさをリアルタイムに表示するシステムを開発し、予備的な評価実験を行った。まだ実験参加者が少なく、断定的なことは言えないが、実験参加者はシステムを使用した場合、システムを使用しない場合と比べ上手く音読ができたという自覚を持つ傾向が見られた。今後の課題として、参加者数を増やすとともに、音響的特徴の変化などの客観的データの分析が挙げられる。また、より高い訓練効果を得るために最適な教示について引き続き検討を行う。

謝辞 本研究はJSPS 科研費基盤研究(A)「ポップアウト・ボイスの生成・知覚基盤の解明に基づく高性能拡声音技術の開発」(JP20H00291)の助成を受けた。

参考文献

- [1] 北村ら, 音響誌, 75(3), 118–124 (2019).
- [2] 鴨居, 沼田, 日本音声学会全国大会予稿集, 127–132 (2022).
- [3] 三谷ら, 音講論(秋), 389–390 (2017).
- [4] Tang *et al.*, *Speech Commun.*, 75, 1–13, (2015).
- [5] 北村ら, 音講論(秋), 1357–1358 (2019).