

口角の高さを指標にした発話訓練法の検討

設楽 郁巳(甲南大学大学院)・安田 奈央・北村 達也(甲南大学)・
牧野 桃子・山根 典子(広島大学)
m2224002@s.konan-u.ac.jp

1. はじめに

医学的には健常に分類される人の中にも、発話に不自由を感じている人が存在する。北村ら(2019)が国内の15大学の大学生、大学院生を対象に行ったアンケート調査では、全体の31.0%がある程度の発話のしにくさを自覚しており、そのほとんどが発話の改善を希望していることが報告された。発話のしにくさを自覚したままの場合、円滑なコミュニケーションを行うことができず、就職活動などの面接やプレゼンテーションなどの場面において実力を発揮できない可能性がある。鴨居・沼田(2022)は発音、発声、滑舌などを学ぶことで、人前で話すことに対する不安や心配、自信がない等の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。

明瞭性が高く聴き取りやすい発話を行うために、様々な発話訓練法が検討されている。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が、その効果や理由は明確に示されていない(三谷ら, 2017)。Clear speech 発話時の顔面の動きは Plain speech 発話時より大きい(Tang ら, 2015)ことに基づいて、北村ら(2019)は発話時の顔のオプティカルフローの大きさをリアルタイムにフィードバックする発話訓練支援システムを提案した。しかし、このシステムは計算量が大きいと動作が高速とは言えず、移植性も低い。この問題点を克服した発話訓練支援システムの開発も試みられたが、予備的な評価実験では十分な訓練効果は確認されなかった(設楽ら, 2022)。

一方で、別の発話訓練法として口角を上げて笑顔で発話するよう促す指示が用いられることもある。Keough ら(2015)は笑顔が調音に及ぼす影響として、母音/i/, /a/, /u/のすべてで唇の広がりや口角の挙上、基本周波数の上昇が見られ、/a/ではフォルマント周波数 F1, F2 の上昇も確認されたことを報告した。

本研究では、アナウンスの発話訓練では口角を上げるよう指導される点に着目した。発話時の口角の位置を上昇させることで発話が改善されるという仮説を立て、口角の高さを指標として用いる発話訓練法の検討を行った。

2. 発話時の口角位置調査

2.1. 目的

発話時の顔の動きと発話訓練経験との関連を確認するため、発話時の顔画像を撮影し、発話時間全体のうちどれだけの間口角が上昇していたか調査した。本実験にて使用した映像データは、安田・北村(2023)にて収録されたものの一部である。

2.2. 実験参加者

大学の放送サークルに所属し発話訓練経験のある大学生 6 名(A 群), 発話訓練経験のない大学生 10 名(B 群), 計 16 名を対象とした。

2.3. 発話資料

前川(2021)を参考に、イソップ童話「北風と太陽」の文章を使用した。使用箇所は以下の通りである。

ある時、北風と太陽が力くらべをしました。旅人のマントを脱がせた方が勝ちということに決めて、まず北風から始めました。旅人は、北風が吹けば吹くほどマントをしっかりとからだにくっつけました。今度は太陽の番になりました。太陽は雲の間から暖かな光を送りました。旅人は段々よい心持ちになって、しまいにはマントを脱ぎました。そこで北風の負けになりました。

2.4. 実験手順

収録に先立ち、書面により実験説明を行い、署名による同意を得た。

話者は椅子に座り、タブレットアームで固定したタブレットに表示された発話資料を読み上げた。話者の顔の正面から約 0.4 m の位置にビデオカメラ(Panasonic HC-V360MS)を固定し、画像サイズ 1080x1920 pixel, フレームレート 30 fps にて話者が発話資料を読み上げている最中の動画を撮影した。撮影時、話者の顔がビデオカメラに対し正面を向くようにするため、ビデオカメラ及びタブレットの高さを話者の座高に合わせて調節した。発話資料の読み上げ方法を、A 群の話者には「滑舌よくアナウンサー風の発話で」と指示し、B 群の話者には「友達と話すような普段通りの発話で」と指示した。

2.5. 口角位置の判定

発話中の口角の動きを確認するため、収録した動画のうち発話区間のみを取り出し調査を行った。口角の位置の判定には、機械学習ライブラリ MediaPipe (Lugaresi ら, 2019)の Face Mesh を使用した。Face Mesh には、目や鼻、口唇といった顔面の特徴的な点である顔ランドマーク座標を推定する機能がある。本研究では、左右の口角と、上唇中央の最下部、下唇中央の最上部の 4 点の座標を抽出した。上唇と下唇を結んだ線の中点の座標を計算し、その中点が左右の口角を結ぶ線の上どちらに存在しているか計測した。図 1, 2 はそれを模式的に表したものである。図 1 のように中点の座標が左右の口角を結ぶ線よりも上であれば、口角の位置は下がっているものと判定される。図 2 のように中点の座標が左右の口角を結ぶ線よりも下であれば、口角の位置は上がっているものと判定される。この判定法を用いて、2 群の話者が発話時間全体の内、どれだけの間口角が上昇していたか調査した。発話時間全体に対する口角が上位であった時間率のことを、本研究では口角上位時間率と呼ぶ。

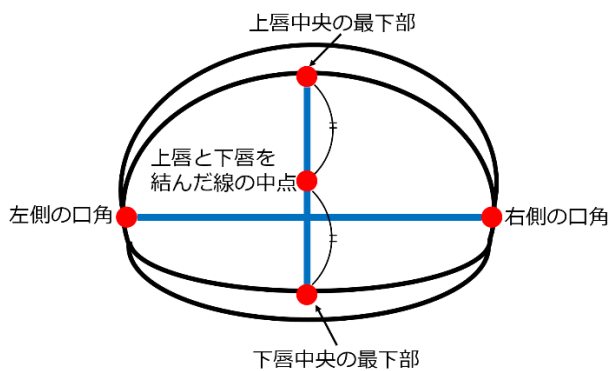


図 1：口角が下の状態

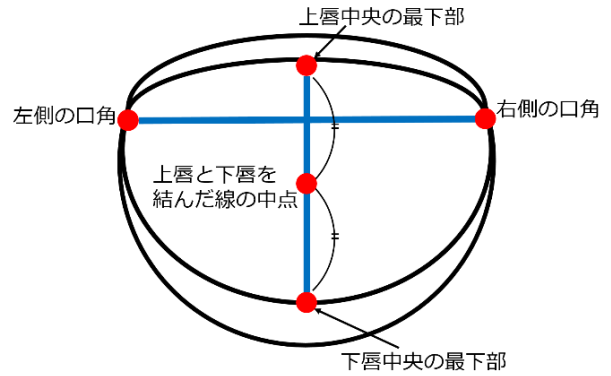


図 2：口角が上の状態

2.6. 結果

調査結果を図 3 に示す。A 群(発話訓練経験あり)では、平均して発話時間全体の 83.6%の間口角が上に位置していた。それに対し、B 群(発話訓練経験なし)では平均して 12.8%の間しか口角が上がっていなかった。

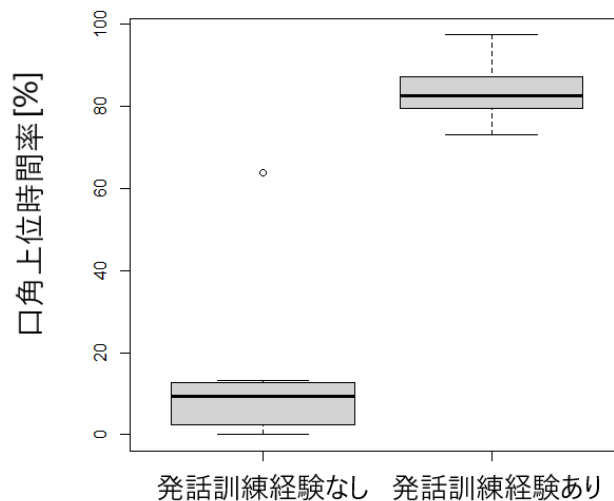


図 3：両群の発話時における口角上位時間率

2.7. 仮説

以上の調査から、A 群と、B 群の発話中における口角上位時間率には大きな差があることが分かった。そこで、発話時の口角の位置を上昇させることが、発話の改善に繋がるという仮説を立てた。この仮説を検証するため、口角の位置を指標として用いる発話訓練法の検討を行い、予備的な評価実験を実施した。

3. 口角の高さを指標にした発話訓練法の検討

本研究では、発話時の口角の位置を上昇させることで発話が改善するという仮説をもと

に、口角の高さを指標として用いる発話訓練支援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4に示す。システム利用者はPCのWebカメラに自分の顔を映し、口角の位置を確認しながら発話を行う。口角の高度判定には2.5節のアルゴリズムを用いた。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、それに重なる形で左右の口角とそれを結ぶ線、上唇と下唇の中点の位置が表示される。画面中央に四角形の枠が表示され、利用者はこの枠内に自分の顔が収まるよう意識する。利用者の口角の位置が下の場合、画面には「口角が下です」と表示し、枠の色を青にする。口角の位置が上の場合、画面には「口角が上です」と表示し、枠の色を赤にする。利用者はこの表示をできるだけ長い時間維持しながら発話を行う。画面上部には発話練習の残り時間と練習文が表示され、練習文はあらかじめ設定した時間ごとに切り替わる。

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。

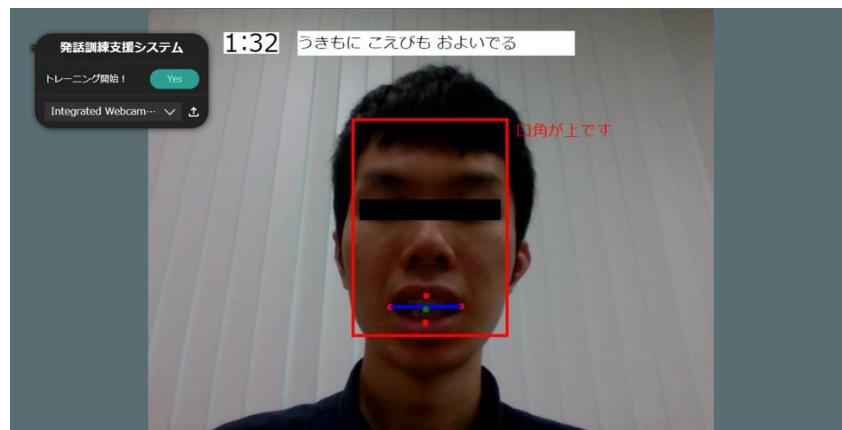


図4：提案システムのスクリーンショット

4. 実験

4.1. 実験参加者

日本語を母語とする18歳から24歳の大学生・大学院生の男女計20人(男性17名、女性3名)が実験に参加した。

4.2. 発話資料

発話資料は北原白秋「あめんぼ」の全文と、2.3節に示したイソップ童話「北風と太陽」の文章を使用した。

4.3. 実験手順

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、本研究で検討した発話訓練法を使用するA群10名と、発話訓練法を使用しないB群10名の2群に分けた。

実験参加者は椅子に座り、ディスプレイ(BenQ GW2270-T)に表示された発話資料を読み上げた。話者の顔の正面から約0.5mの位置にウェブカメラ(Logitech BRIO)を配置し、画像サ

イズ 1280x720 pixel, フレームレート 30 fps にて話者が発話資料を読み上げている最中の動画を撮影した。ビデオカメラ及びディスプレイの高さは実験参加者の座高に合わせ調節した。発話資料の読み上げの後、自分の発話がどれくらい上手く読めたか Visual Analogue Scale (VAS)にて自己評価させた。動画とは別に音声の収録も行った。コンデンサマイク (Sony ECM-77B), オーディオインターフェース(Roland Rubix24)を用いて標本化周波数 44.1 kHz, 量子化ビット数 24 bit にて録音した。マイクはマイクスタンドを用いて、話者頭部の正中から左に約 30° の地点に配置し、口唇とマイクとの距離は約 0.25 m とした。

次に、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 で接続し、自己練習を行うよう指示した。両群の自己練習ではまず、(1) 準備体操動画, (2) 顔マッサージ動画の順番でデモンストレーション動画を視聴させた。A 群ではこれに加え(3)口角を上げるためのデモンストレーション動画の視聴に続き、発話訓練支援システム(図 4)を使用させ、1 分 40 秒の発話訓練を 5 回実施した。一方 B 群には、PC の画面中央に表示した文が一定間隔ごとに切り替わる別のプログラムを使用させた。この手続きを 1 日 1 回、2 週間の間実施した。1 回に要する時間は動画の視聴も合わせ約 15 分程度である。

自己練習開始から 1 週間及び 2 週間経過時に、実験参加者の顔画像の撮影と、音声の収録を再度行った。撮影、収録の条件は同一である。ただし、A 群には発話資料の読み上げ方法を、「発話訓練支援システムで練習したように、口角を上げるよう意識して」と指示した。

5. 結果と考察

A 群、B 群の「北風と太陽」発話時の、口角上位時間率(%)の遷移をそれぞれ図 5、図 6 に表す。横軸は実験開始から経過した日数を表している。A 群では 7 日目以降の収録で口角上位時間率が向上する傾向が見られたのに対し、B 群では初回の収録時と 2 回目以降もほぼ変わらない傾向が見られた。収録した音声に対し、音響的特徴の変化が見られるか解析中である。著者の聴覚印象としては、口角上位時間率の向上が見られた実験参加者の発話は、訓練前と比べ明瞭性が改善されたように感じられた。

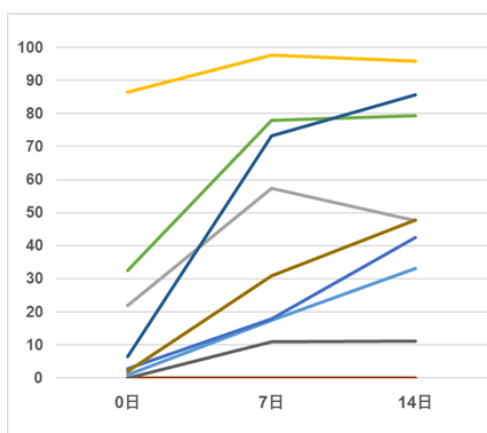


図 5 : A 群の口角上位時間率

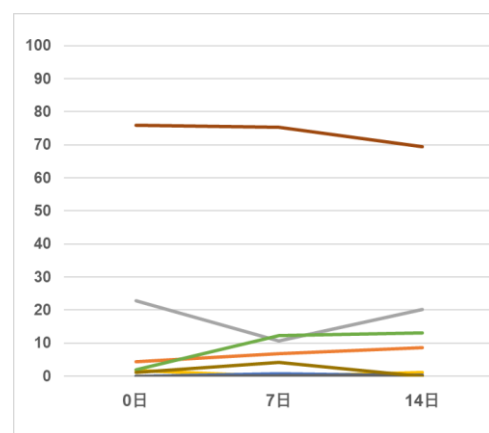


図 6 : B 群の口角上位時間率

6. まとめ

本研究では、発話時の口角の位置を上昇させることで発話が改善するという仮説をもとに、発話訓練法の検討を行った。予備的な評価実験を行った結果、検討した発話訓練法を用いた群では、文書読み上げ中の口角上位時間率が向上する傾向が見られた。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。口角上位時間率と音響的特徴との相関も次のステップとして考察することになるであろう。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予定である。

謝辞 収集データ整理にご協力いただいた甲南大学知能情報学部 渡邊ひかりさんに感謝します。本研究は、中山隼雄科学技術文化財団, 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) の助成を受けています。

参考文献

- 鴨居真理子・沼田秀穂 (2022)「日本語音声教育の効果の検証 -アナウンススクールの事例-」『日本音声学会全国大会予稿集』 127-132.
- Keough, M., A. Ozburn, E. McClay, M. Schwan, M. Schellenberg, S. Akinbo, and B. Gick (2015). “Acoustic and articulatory qualities of smiled speech”, *Canadian Acoustics*, 43(3), 126-127.
- 北村達也・川村直子・能田由紀子・吐師道子 (2019)「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」『日本音響学会秋季講演論文集』 1357-1358.
- 北村達也・能田由紀子・吐師道子・竹本浩典 (2019)「大学生・大学院生を対象とした発話のしにくさの自覚に関するアンケート調査」『日本音響学会誌』 75(3), 118-124.
- Lugaresi, C., J. Tang, H. Nash, C. McClanahan, E. Uboweja, M. Hays, F. Zhang, C. Chang, M. Yong, J. Lee, W. Chang, W. Hua, M. Georg, and M. Grundmann (2019). “MediaPipe: A Framework for Building Perception Pipelines”, arXiv:1906.08172.
- 前川喜久雄 (2021)「リアルタイム MRI 動画データベースプロジェクトの概要と調音音声学への応用」『言語資源活用ワークショップ 2021 発表論文集』 96-107.
- 三谷巧・尾下克樹・北村達也・川村直子・能田由紀子・吐師道子 (2017)「開口制約に基づく発話訓練法の効果」『日本音響学会秋季講演論文集』 389-390.
- 設楽郁巳・関和広・北村達也 (2023)「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」『日本音響学会春季研究発表会』 583-584.
- Tang, L., B. Hannah, A. Jongman, J. Sereno, Y. Wang, and G. Hamarneh (2015). “Examining visible articulatory features in clear and plain speech”, *Speech Communication*, 75, 1-13.
- 安田奈央・北村達也(2023)「発話訓練経験による文章発話時の顔ランドマーク変位の違い」『日本音響学会春季研究発表会』 613-614.