日本音響学会第149回研究発表会
2023/03/16

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

甲南大学知能情報学部
☆安田奈央，　北村達也

明瞭な発話と顔の動きの関係

明瞭な発話がもたらす影響

- 発話のしにくさの緩和
- 円滑なコミュニケーション

発話と調音運動の関連

- 発話スタイルによる唇の伸張や突出、顎の変位の違い(Tangら 2015)
- 明瞭度の異なる音声発話時の舌と顔の動きの相関(Jiangら2002)

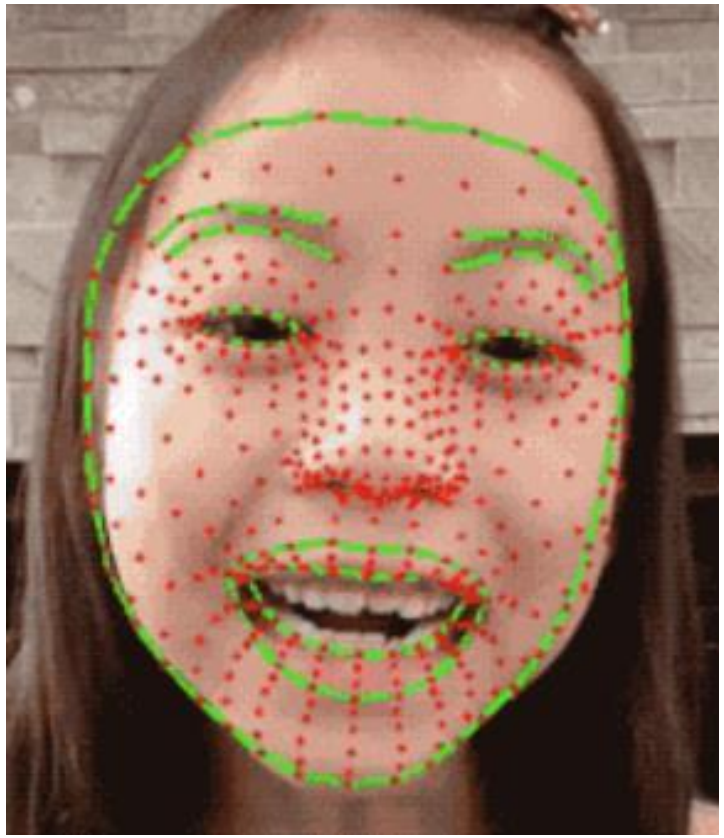
音声の明瞭性が高い人ほど顔の動きが大きい？
顔の動きを発話訓練へ利用

顔ランドマーク座標の取得方法

Face Mesh

- 画像認識クラウドAIシステム
- 顔ランドマーク468点の3次元座標推定
- 変動性や重要性の高い部分に
  ランドマークが多く割り当てられる

発話中の顔画像を撮影
フレームごとの顔ランドマーク座標を取得



方法１ データ収録（1）

●実験参加者　18名

A群	プロのナレーター	2名　（男１，女１）
B群	大学生(放送サークル所属)	6名　（男１，女５）
C群	大学生(発話訓練経験なし)	10名（男５，女５）

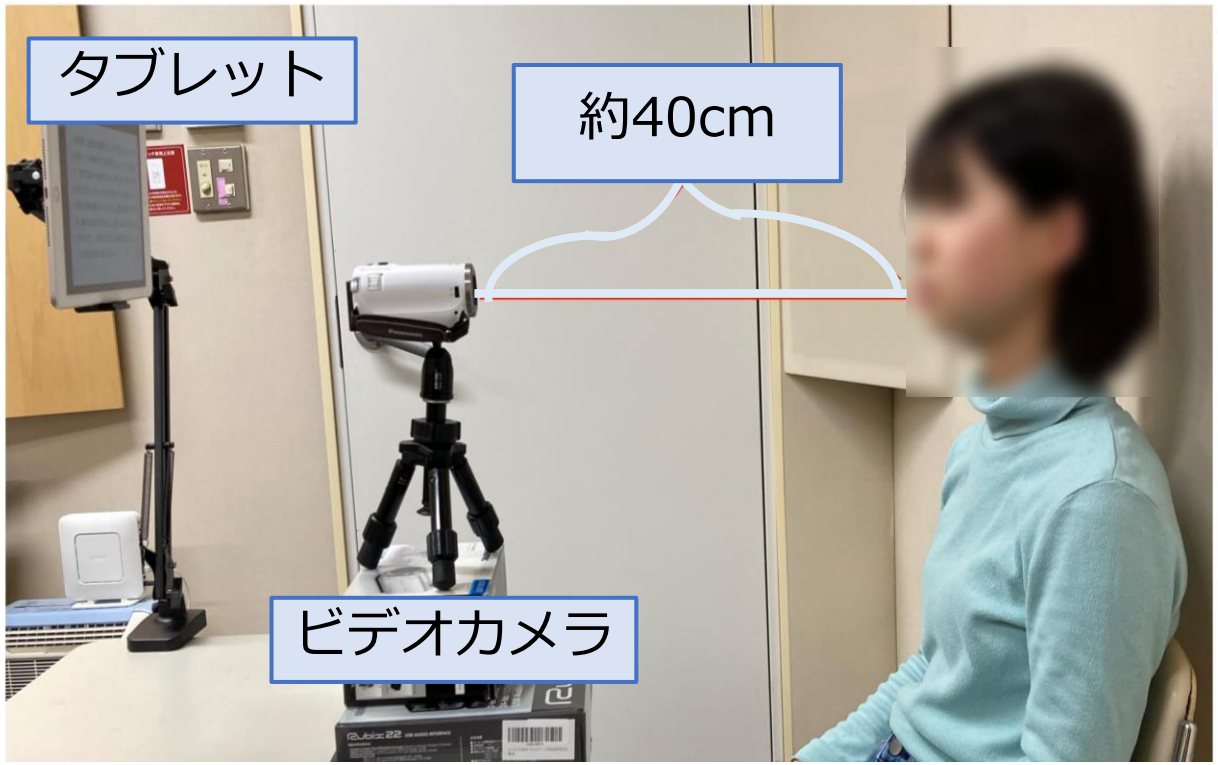
●発話資料　「北風と太陽」

A群	滑舌よくアナウンサー風の発話で
B群	滑舌よくアナウンサー風の発話で
C群	友達と話すような普段通りの発話で

方法１ データ収録（2）

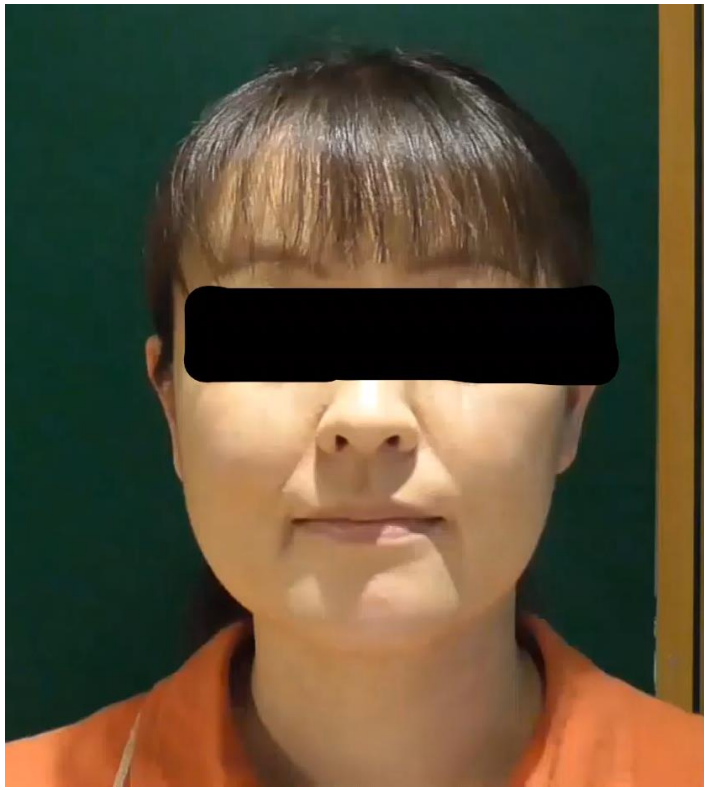
●発話中の顔画像を撮影

- ビデオカメラ (Panasonic HC-V360MS)
- 画像サイズ：1080×1920 pixel
- フレームレート：30 fps

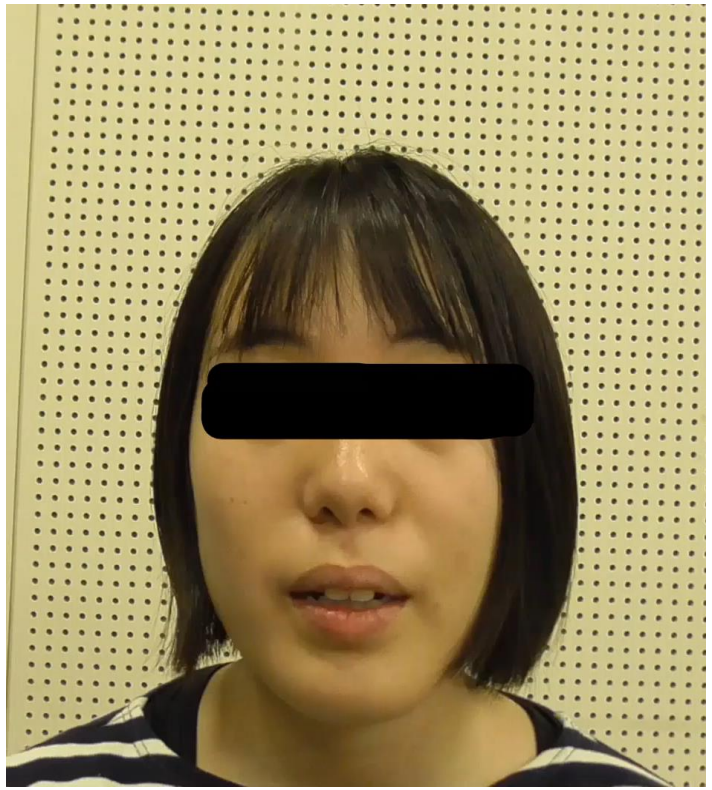


撮影した動画の例

A群　女性



B群　女性



C群　女性

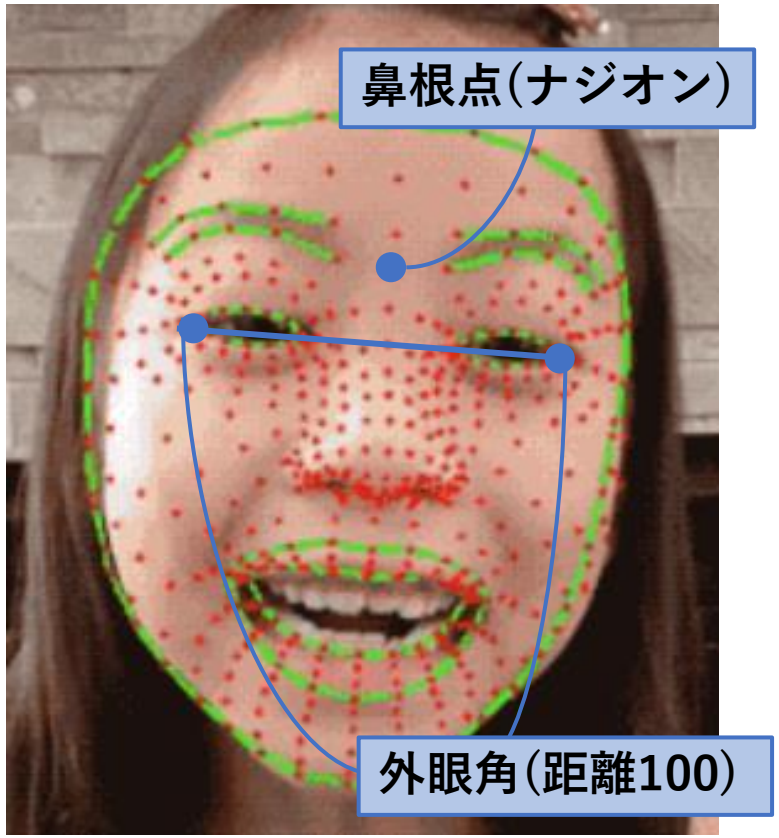
方法2 顔ランドマークの分析（1）

フレーム間の移動距離

- 各点が1フレーム間に移動した距離を算出

個人差に対する処理

- 頭部の動き：鼻根点（ナジオン）を基準に
- 顔のサイズ：外眼角の距離を相対値に

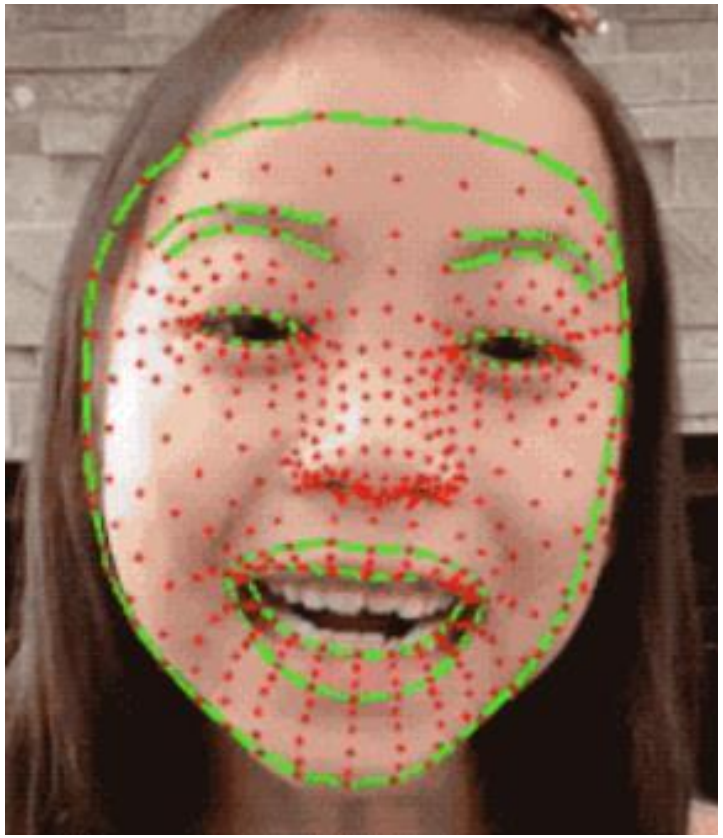


方法2 顔ランドマークの分析（2）

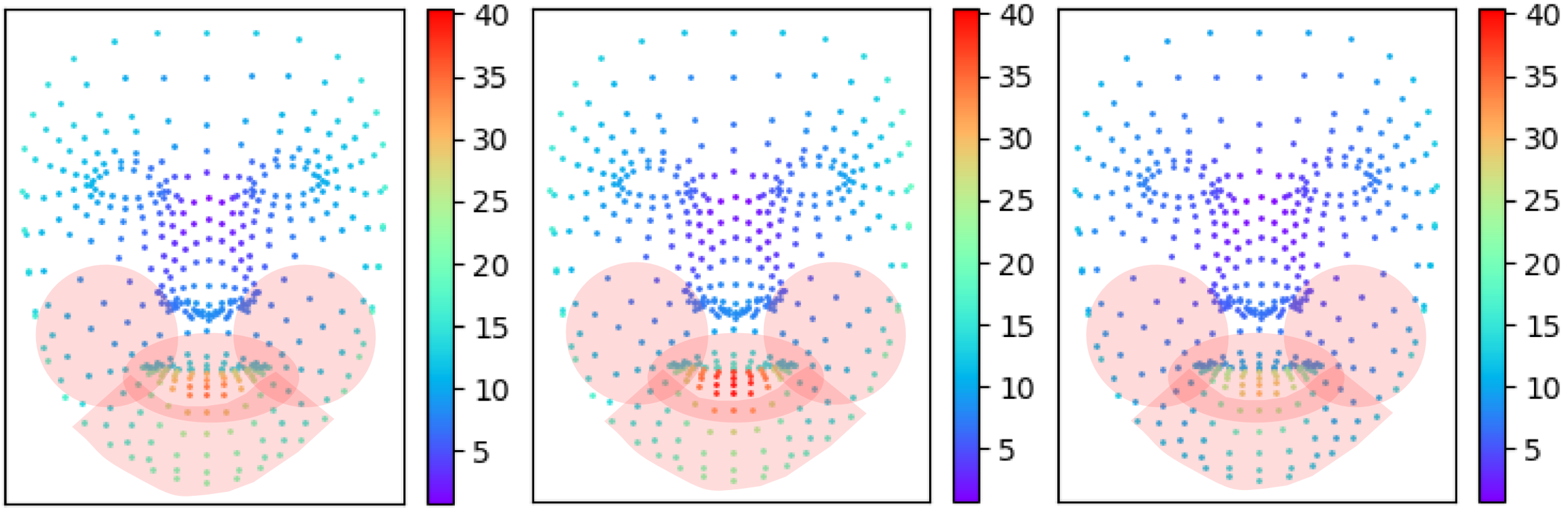
発話中の１sあたりの
各顔ランドマーク移動量を算出

- 全フレーム間の移動距離を合計
- 文章読み上げ時間で割る

話者群ごとに平均移動量を算出・比較
B群とC群の間で顔の動きの有意差検定



結果1 各顔ランドマークの平均移動量



(a) A群
プロのナレーター

(b) B群
放送サークル所属

(c) C群
発話訓練経験なし

結果2 B群とC群間の有意差

有意差検定

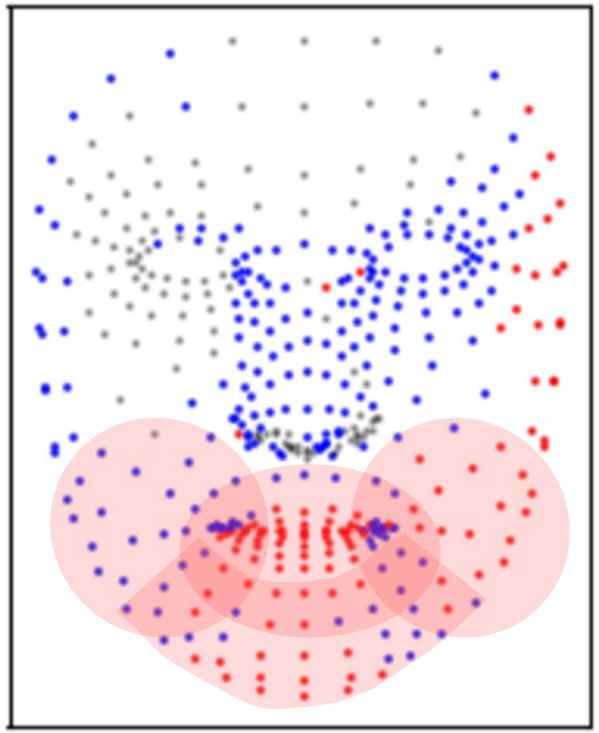
- Mann-Whitney U検定を用いる

有意差あり

- 口唇周辺
- 下顎
- 頬の領域

※鼻梁、顔側面でも動きに有意差あり

鼻梁の変位量、撮影状態、一部の話者の動きが影響



赤：1%水準で有意.
青：5%水準で有意. 灰色：非有意

結果3 Z軸方向の平均変位

Z軸方向の平均変位を群間で比較

- A群、B群、C群の順に大きい
- 平均変位量は極めて小さい

有用な情報は得られなかった

横方向からも撮影する必要あり？

考察とまとめ

発話訓練経験の異なる話者の顔ランドマークの動きを比較

- Face Meshを用い、高速に顔ランドマーク座標を取得・分析
- 発話訓練の有無で口唇、下顎、頬の動きに差異・有意差あり
- 音声が明瞭な話者群は顔の動きが大きい

顔の動きを発話訓練へ利用

- 手本となる顔ランドマークの動きを発話訓練へ利用できないか？
- 顔の動きや音声の明瞭度に変化？

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・近く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。
収録にご協力いただいた広島大学 山根典子先生、牧野桃子様に感謝します。

音声録音

●収録機材
- コンデンサマイク（RODE NT2-A）
- オーディオインタフェース（Roland Rubix24）

●収録方法
- 標本化周波数　：48 kHz
- 量子化ビット数：24 bit

検定方法の決定

●対応のない2群の検定

●正規性：シャピロ・ウィルク検定
●等分散性：F検定

●正規性あり/等分散性あり：Student t 検定
●正規性あり/等分散性なし：Welch’s t 検定
●正規性なし/等分散性なし：Mann-Whitney U検定