6V-09

Web 会議における発話衝突抑制のための顔特徴を用いた発話予測手法の提案

山田　楓也†　　白石　陽†
公立はこだて未来大学システム情報科学部†

1. はじめに

近年，テレワークの普及に伴って Web 会議サービスの利用が増加している．Web 会議では，画面サイズが小さいことから，相手の状況や会話時の様子が読み取りにくく，発言のタイミングを掴みにくい．その影響により，発言者同士の発話の重なり（発話衝突）が発生することがある．発話衝突は，参加者が会話終了後に再発話する際に，同じタイミングで複数の参加者が発話した場合に発生する．発話衝突が多発することで，会話の中断や会議参加者の発言意欲の低下につながり，消極的な会議に発展する可能性がある．そこで，発話衝突を抑制することができれば，円滑な会議の進行が期待できると考える．

本研究では，発話衝突を抑制するために，発話予測を行う．対面会議において，会議参加者は他の参加者に視線を向ける，顔を傾けるなどといった発話前に行う特徴的な動作（予備動作）を行うことから，発話と顔情報に関係性があると考えられる．そこで，顔情報に着目することで発話タイミングを予測し，他の参加者へ予測情報を提示する．これにより，参加者がいつ発話するべきか，または他の参加者に発話を促すかといったことが可能になり発話衝突を抑制できると考える．

本研究では，Web 会議における顔特徴を用いた発話予測手法を提案する．本稿では，発話予測に有効な顔特徴を調査するために，個人ごとの機械学習モデルを構築し，顔特徴ごとの精度評価の比較を行った．

2. 関連研究

発話予測の関連研究として，石井ら[1], [2]の研究と玉木ら[3]の研究がある．

石井らは，複数人による対面での会話を想定し，次に発話する話者（次話者）の予測を行っている[1], [2]．文献[1]では頭部運動に着目しており，文献[2]では視線交差のタイミング構造に着目している．しかし，Web 会議を想定した場合，視線移動や頭部運動の特徴は対面での会話に現れる動作と異なるため，これらの手法は本研究に適用できないと考える．

玉木らは，複数人による Web 会議での会話を想定して，予備動作を用いた発話予測を行っている[3]．Kinect センサを用いて手を挙げる，頷く，手を顔に近づけるの予備動作を検出し，検出した動作をインジケータで発話欲求度合いを参加者に提示するシステムの構築を行っている．発話欲求度合いは，予備動作を検出するごとにスコアを加算し，その総和としている．しかし，文献[3]の予備動作は個人差を考慮していない．そのため，上記の予備動作以外の予備動作を検討する必要がある．

A Proposal of a Method for Speech Prediction Using Facial Features to suppressing speech Contention in Web Meetings
†Fuya Yamada †Yoh Shiraishi
†School of Systems Information Science, Future University Hakodate

3. 提案手法

3.1 研究目的

本研究の目的は，Web 会議における発話衝突を抑制するために，顔特徴を用いて発話予測を行うことである．

3.2 提案システム

本研究の提案システムの全体像を図 1 に示す．

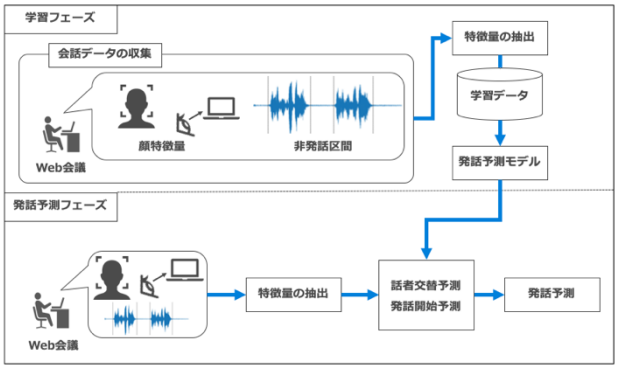

図 1 提案システムの全体像

提案システムは，学習フェーズと発話予測フェーズの 2 つのフェーズで構成される．学習フェーズは，Web 会議サービスを用いた会話時の映像データから，発話者が発話する前の無音時間（非発話区間）における参加者の顔特徴点の位置変化（顔特徴データ）を収集し，発話予測に用いる特徴量を抽出する．抽出した特徴量を正解ラベルとともに学習データとする．学習データを用いて発話予測モデルを構築する．発話予測フェーズでは，Web 会議から収集した映像データから，特徴量を抽出し，構築した発話予測モデルを用いて発話予測を行う．

3.3 発話予測の流れ

発話予測の流れとして，(1)顔特徴データの収集，(2)特徴量の抽出，(3)発話予測の手順で発話予測を行う．

(1)では，Web 会議サービス利用時における顔特徴データの収集を行う．映像データは表 1 に示す顔特徴データに変換する．

表 1 発話予測に用いる顔特徴データ

種類	内容
視線移動	水平方向 (gaze_angle_x) 垂直方向 (gaze_angle_y)
頭部運動	水平方向 (pose_Tx) 垂直方向 (pose_Ty) カメラとの距離 (pose_Tz)
口の開き	Landmark から口の開き具合を算出

複数人で行う会議を想定して，収集した映像データから顔特徴データを抽出する．顔特徴データの分析区間は，

非発話区間とする．非発話区間は，注釈ツールである ELAN を用いて発話終了から次の発話の直前を注釈する．この非発話区間から顔特徴データを抽出する．

(2)では，(1)で抽出した顔特徴データから，特徴量を抽出する．顔特徴データから基本統計量（平均値，標準偏差，最小値，最大値，中央値）5 種類を計算し，特徴量として抽出する．

(3)では，(2)で抽出した特徴量と発話予測モデルを用いて話者交替予測と発話開始予測を行う．識別器は，話者交替予測には SVM (Support Vector Machine) を用いて，発話開始予測には SVR (Support Vector Regression) を用いる．

4. 実験および考察

4.1 実験目的

本実験では，発話予測に有効な顔特徴を調査するために，個人ごとの発話予測モデルを構築し，顔特徴データの組み合わせごとのモデルの精度評価を比較した．

4.2 実験環境

実験の被験者は，20 代の男子大学生 3 名とした．被験者全員が Web 会議サービスの Zoom を使用し，10 分程度のファシリテーションを設けないアイディア出しを行った．画面全体の映像，全体の合成音声，各参加者の音声と映像データの収集を行った．収集した映像データから OpenFace[4]を用いて顔特徴データの収集を行った．また，SVM を用いて，発話予測モデルの構築を行った．

4.3 収集した顔特徴データの分析

非発話区間における顔特徴データを収集したデータの一例として，被験者 A の視線方向のデータをグラフ化したものを図 2 に示す．縦軸は OpenFace による出力値，横軸は時系列データのサンプル番号を表す．また，青の領域は，非発話区間を表す．

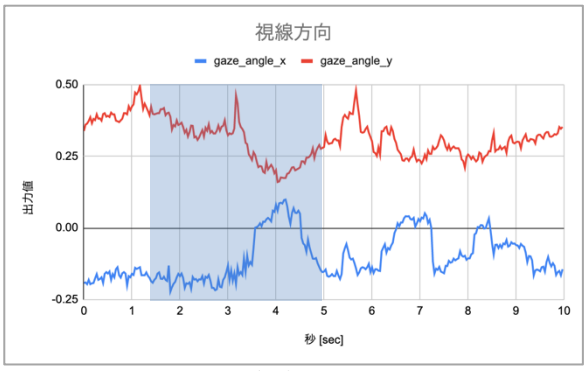

図 2 視線データの一例

視線移動の水平方向 (gaze_angle_x) と垂直方向 (gaze_angle_y) は，3 秒から 5 秒にかけて大きく変動していることがわかる．画面に写っている参加者を見るために，視線を移動したことが推測される．そのため，非発話区間において特徴的な傾向が現れると考えられる．

4.4 評価結果と考察

本実験では，発話予測に有効な顔特徴を調査するため，話者交替と話者継続の 2 クラスに分類した．精度評価として，10 分割交差検証で評価した際の F 値を使用した．

表 1 の顔特徴データから基本統計量を算出し，発話予測モデルを構築した．構築したモデルを用いて発話予測を行った．その予測結果を表 2 に示す．

表 2 発話予測の評価結果

モデルに使用した顔特徴	F 値
視線移動のみ	0.553
頭部運動のみ	0.599
口の開きのみ	0.508
視線移動＋口の開き	0.456
視線移動＋頭部運動	0.631
頭部運動＋口の開き	0.583
視線移動＋頭部運動＋口の開き	0.549

表 2 より，視線運動と頭部運動の組み合わせの場合において F 値の値が 0.631 となった．この結果から，被験者 A 名の予備動作として視線移動と頭部運動を用いることが有効であることがわかった．しかし，今回構築したモデルの精度は不十分であると考える．そのため，今回の実験に使用した顔特徴データ以外にも使用する顔特徴データの検討を行う必要があると考える．

5. おわりに

本研究の目的は，Web 会議における発話衝突を抑制するために，顔特徴データを用いて発話を予測することである．本稿では，Web 会議中に Web カメラから顔映像と音声を収集し，顔情報から抽出した特徴量を用いて発話予測を行った．その結果，視線移動と頭部運動を場合の F 値が 0.631 で最大となった．本稿で述べた実験の被験者は予備動作として視線移動と頭部運動に関連性があることがわかった．

今後は実験で採用した特徴量をもとに発話開始予測を行う．さらに，予測モデルの精度向上のために，他の顔特徴の組み合わせも検討する．また，個人の発話予測だけでなく，参加者全員の発話予測も行う予定である．

参考文献

[1] 石井 亮，大塚 和弘，熊野 史朗，大和 淳司，“複数人対話における頭部運動に基づく次話者の予測”，情報処理学会論文誌，Vol.57，pp.1116-1127 (2016).

[2] 石井 亮，大塚 和弘，熊野 史朗，大和 淳司，“複数人対話における視線交差のタイミング構造に基づく次話者と発話開始タイミングの予測”，人工知能学会全国大会論文集，Vol.29，pp.1-4 (2015).

[3] 玉木 秀和，東野 豪，小林 稔，井原 雅行，“発話がぶつからない Web 会議を実現するための発話欲求伝達手法”，情報処理学会論文誌，Vol.54，No.1，pp.275-283 (2013).

[4] T.Baltrusaitis，P.Robinson，and L-P.Morency，“OpenFace: An Open Source Facial Behavior Analysis toolkit,” Proc. of the 2016 IEEE Winter Conference on Applications of Computer Vision，pp.1-10 (2016).