マルチモーダル情報の提示を含んだ自動ソーシャルスキルトレーニングの訓練効果*

◎田中宏季, サクリアニ・サクティ，グラム・ニュービッグ（奈良先端大），
根來秀樹，岩坂英巳（奈良教育大），中村哲 (奈良先端大)

1 はじめに

ソーシャルスキルトレーニング（SST）は幅広く社会的コミュニケーションを苦手としている人々に適用されている訓練手法であり， 医療機関，学校などで人間のトレーナにより実施されている．SST の全体もしくは一部分をコンピュータで自動化することができると，希望者がいつでもどこでも、SST を受けることができると考えられる．そこで，我々は音声対話システムを用いて SST の自動化を行う研究を進めており， コンピュータを用いて従来の SST を模倣した「自動ソーシャルスキルトレーナ」を提案し，SST としての有効性を確認した [1].

しかしながら，これまでの自動ソーシャルスキルトレーナは音声および言語情報のみしか考慮していないという問題があった．実際の SST では，表情や姿勢など視覚情報も含めたフィードバックを行っていることから [2]，これらのマルチモーダルな情報も組み込む必要があると考えられる．画像処理を中心とした面接訓練の研究も存在していることから [3]，本稿では既存の SST の枠組みに従い，画像情報を含めたフィードバック提示システムの構築を行った．実験により，画像情報がどの程度 SST として有効か調査を行った．

2 SST とマルチモーダル情報の提示

SST には様々なモジュールがあるが，本研究ではそのうち基本訓練モデルを採用した [2]．SST の基本訓練モデルは，課題設定，モデリング，ロールプレイ，フィードバック，正の強化，宿題の各モジュールにより構成される．課題設定の例としては，4 つの基本的なスキル [2] が広く使用されており，それぞれ「うれしい気持ちを伝える」,「頼み事をする」,「相手の言う事に耳を傾ける」,「不愉快な気持ちを伝える」となっている．我々は自動ソーシャルスキルトレーナを開発し，課題として「うれしい気持ちを伝える」を選択している [1]．本節では，当課題についてロールプレイとフィードバックに関してのマルチモーダル化について述べる．

2.1 ロールプレイ

ロールプレイでは，ユーザがこれまであった楽しかった話を 1 分間でコンピュータのアバターに伝える．その際，ユーザの音声，言語に加え，画像特徴量をマイクとカメラにより抽出する．抽出する全特徴量について以下にまとめる：1) F0 の変動係数：100Hz 以上の F0 に関する変動係数，2) パワー：パワー値の平均，3) 声質：スペクトル傾斜について，第一倍音と第三フォルマントの差の特徴量，4) ポーズ：ユーザの発話開始までの時間，5) WPM：ユーザが 1 分間発話をするため，その間の単語数，6) 6 文字以上の単語割合：全発話から 6 文字以上の単語を使用していた割合，7) フィラーの割合：「えー」や「ああ」などのフィラーの割合，8) 笑顔の頻度：全フレームから笑顔の割合，9) 横を向く回数：顔の横回転の絶対値の平均，10) 下を向く頻度：顔の縦回転の平均値．

この内,「笑顔の頻度」と「横・下を向く頻度」を新たに画像特徴量として抽出している．

2.2 フィードバック

ロールプレイで抽出した特長量により，ユーザのスキルに関してフィードバックを提示する（図 1）．フィードバックは以下のものを含んでいる：1) ユーザの動画：ユーザは録画された自身の音声と動画を視聴することができる，2) 話のスコア：システムは重回帰モデルにより予測したスコアを表示する，3) モデルとの比較：Z 値によって，現在の話し方により抽出された各特徴量が上手なモデルの平均的な値と，どの程度ずれているのかを表示する，4) 良かった点および修正点：モデルとのずれから，システムは正のコメントおよび修正のコメントを表示する．

3 実験的評価

画像を含んだフィードバック提示が，SST として有効かを確認するために実験を行った．

3.1 手続き

18 名の大学院生（男性 15 名，女性 3 名）を被験者として募集した．被験者はそれぞれ，音声のみに

*Training Effect of Automated Social Skills Trainer with Multimodal Feedback by Hiroki Tanaka, Sakriani Sakti, Graham Neubig (Nara Institute of Science and Technology), Hideki Negoro, Hidemi Iwasaka (Nara University of Education), Satoshi Nakamura (Nara Institute of Science and Technology)

Fig. 1　フィードバック提示画面.

関するフィードバック提示（男性 6 名，女性 3 名），音声および画像に関するフィードバック提示（男性 9 名），の 2 グループに分けられた．各被験者は，50 分間それぞれのシステムによるトレーニングを受けた．トレーニングは SST の基本訓練モデルに従い，課題の説明，モデリング，ロールプレイ，フィードバック，宿題の順で行われた．

トレーニングの事前と事後で，被験者と面識のある人物に向かって話を伝えている様子をカメラで収録し，スキルの評価を行った．収録した動画に対して，SST を実施している奈良の福祉グループ、障がい児放課後支援事業「ぷろぼのスコラ」の SST トレーナ 1 名が話の全体的なスキルに対して 7 段階で評価を行った．偏りを失くすため，被験者および事前と事後を評価する順番をランダムとした．我々は事前と事後での評価値の変化を算出し，2 グループで Student の t 検定（片側）により有意差検定を行った．

3.2　訓練効果

図 2 に事前と事後の評価値の変化を示す．画像を含めたフィードバック提示が有意に有効であることがわかる（$p = 0.026$, Cohen's $d = 0.98$）．また，7 段階評価での 1 評価値の改善は，先行研究 [1, 3] と比較しても高いスキルの向上効果があることを示している．

4　まとめ

我々は音声対話システムによって従来の SST を模倣する自動ソーシャルスキルトレーナを開発した．本稿では画像情報も含めたシステムを構築した．実験を行い，画像を含めることの有効性を確認した．今後は，SST の基本訓練モデルの各モジュールについて，システムの改良を進めていく．

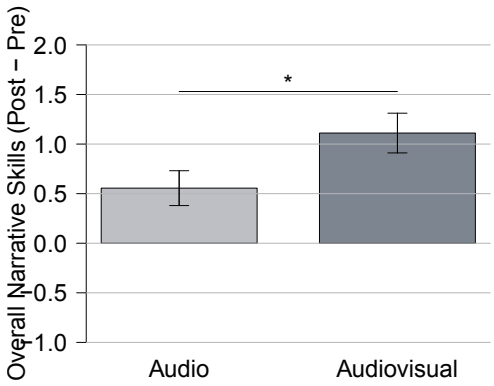

Fig. 2　音声のみと画像を含んだフィードバック提示による評価値の変化 (*: $p < 0.05$)．エラーバーは標準誤差.

謝辞　本研究は，JSPS 科研費 26540117 の助成を受けて行われたものである．

参考文献

[1] Tanaka, H. *et al.* Automated Social Skills Trainer. *Proc. International Conference on Intelligent User Interfaces*, 17-27, 2015.

[2] Bellack, A. S. Social skills training for schizophrenia: A step-by-step guide. *Guilford Press*, 2004.

[3] Hoque, E., *et al.* MACH: my automated conversation coach. *Proc. 15th Conference on UbiComp*, 697-706, 2013.