

習慣省年マンガ

チーム名：ちいばやし

要旨

視点をひっくり返して客観的に省みる漫画アプリ

コミュニケーションにおいて理解不足や齟齬が生じ、気まずい思いをした経験はないだろうか？この気まずさを改善しようとする行動に一人反省会がある。これは自分に非があったという思い込みが原因の行動であり、一人で行うという性質上、「答えの出ない問いを続けて時間を使ってしまう」、「自分の行動や発言の真の影響を見落とす」、「抑うつ気分の悪循環に陥りやすい」などの問題に陥ってしまいがちである。一方で、自分の行動を振り返る習慣がほとんど無い人もいるため、軌轢を助長する1つの要因になっている。

我々はこの問題を解決するため、視点を「ひっくり返す」というコンセプトで、主観視点と客観視点をひっくり返した漫画にすることで省みるアプリである「習慣省年マンガ」を提案する。通常の議事録は主観視点のみだが、このアプリを使うことで客観的な視点による振り返りを促し、思い込みの悪循環を抑制する効果が期待できる。また、普段から言動を振り返る習慣が無い人に対しても、行動変容を促すことができる。

課題と解決策

日常生活において、我々は多くの会話を交わしている。国立国語研究所の調査によれば、1日における平均的な会話回数は12.7回であり、会話に費やされる時間は1日平均で6.2時間に及ぶ[1]。しかし、これほど頻繁に行われるコミュニケーションであっても、他者や社会に対する配慮、さらには双方向の密度の高いコミュニケーションが必ずしも十分に行われているわけではない。理解不足や軌轢が生じやすくなり、日常生活において強い不安や葛藤を感じ、周囲と調和して生活することが困難になる「不適応」を引き起こす傾向がある。この不適応を回避するためには、自己の発言や行動を定期的に振り返ることが求められる。

しかし、この振り返りには2つの問題点がある。1つ目は、一般的に自身のコミュニケーションの振り返りを自発的に行うのは困難である。KPT法[2]やYWT法[3]のような振り返りを効率的に行うための方法が考案されており、これらの方法は、具体的かつ継続的な振り返りを促進することが可能である。しかし、自らが問題点や次回の行動を入力、書き込みを行う必要があり、振り返りの敷居の高さを根本的に解決することは難しい。2つ目は、自身の行動を振り返ることがあっても、しばしば反芻思考としてその人自身の内側で完結してしまい、長時間思考の悪循環が原因で抑うつ状態に陥ってしまう可能性がある[4]。スパイダー討論[5]のような客観的な事実に基づいた振り返りを行うための取り組みは存在するが、話し合いに観察者を参加させて記録を行うなど、人手が必要であり、手軽に実施できない問題がある。

我々が提案する「習慣省年マンガ」は、会話の音声データから自動で自身の行動を客観的な視点にひっくり返した漫画を生成し、提供する。これにより普段振り返りを困難と感じ、振り返りを行っていない人でも簡単に自身の言動を振り返ることができる。また、一人で思い悩み続けてしまう人に対しても、客観的な視点による振り返りを促進し、反芻思考の悪循環を抑制することができる。

図1に習慣省年マンガの利用イメージを示す。習慣省年マンガは会話の音声を録音することで漫画が生成される。その結果、ユーザは自分自身の行動や発言を客観的かつ視覚的で親しみやすい漫画形式で振り返ることができる。

1.会話を録音



2.通知



3.漫画を読む



4.省みる



図1 習慣省年マンガの利用イメージ

習慣省年マンガの機能

「習慣省年マンガ」は、①会話内容の録音、②録音内容の文字起こしと客観的ストーリー、助言文章を生成、③文字起こしした文章から漫画の1コマ目の生成、④1コマ目以外のコマの生成の4つの処理を行う。

1コマ目に生成したキャラ画像をもとに2コマ目以降のマンガを生成するので、先に1コマ目を出力する必要がある。自身の行動を客観的に評価し、改善点を教えてくれる助言文章も同時に生成する。図2に一連の処理の流れを示す。

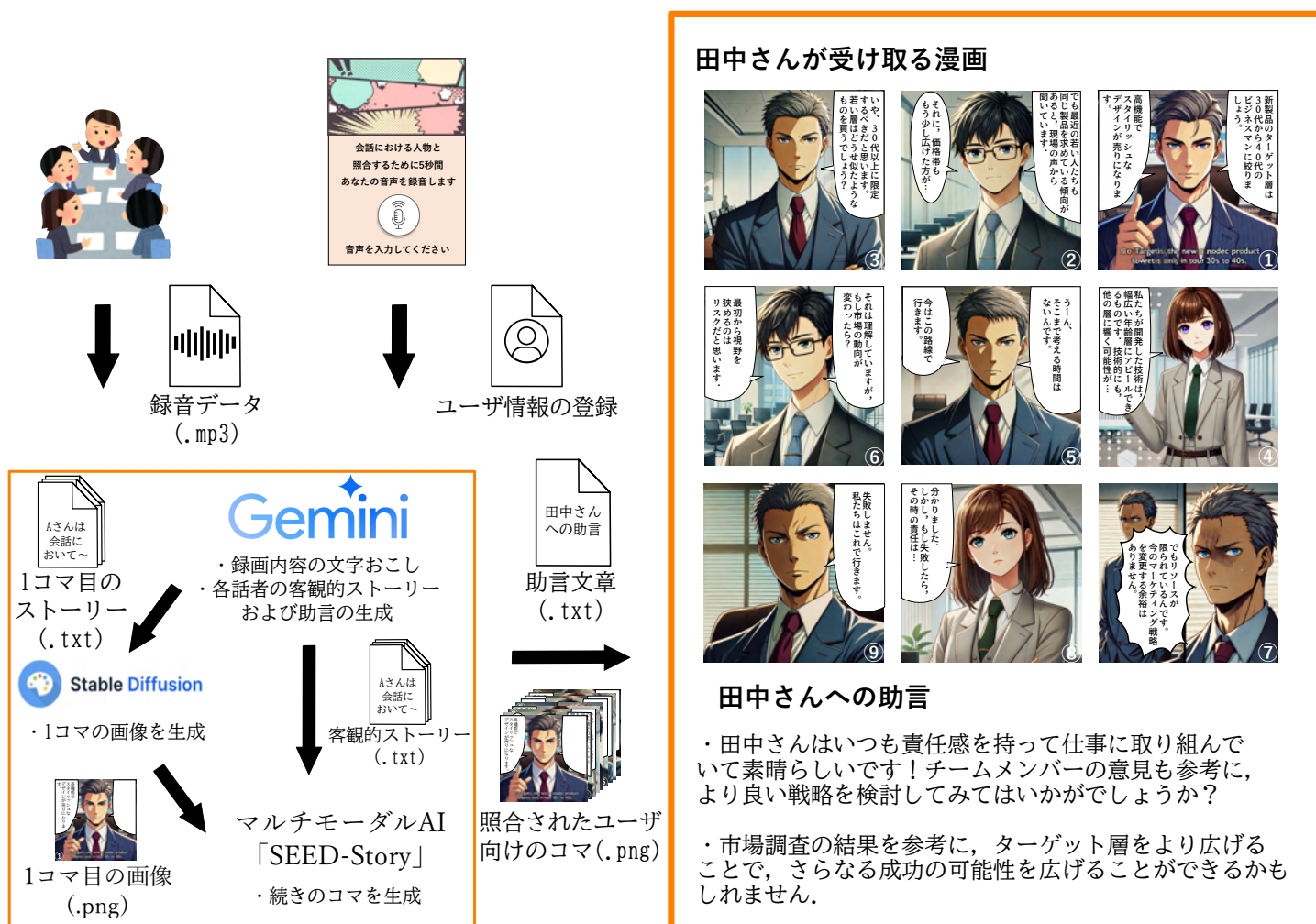


図2 システムの流れ

各機能の詳細について以下に示す。

● ユーザ情報の登録

ユーザに音声入力を促し、名前と音声の情報を紐づけて登録する。

● 録音

会話を始める際に録音を開始し、会話が終了した際に録音を終了する。録音した会話はmp3形式の音声ファイルとして出力する。録音には、ボイスレコーダーやPCマイクなど様々なデバイスで録音可能である。

● 文字起こし

図3に文字起こしの流れを示す。文字起こし中の処理の流れとして、①録音ファイルの文字起こしと発言者の分類、②性別の特定、③ストーリー文章と助言の生成の3つの処理を行う。

①録音ファイルの文字起こしと発言者の分類、発話開始時間の取得

Google Gemini のAPIを用いて、録音ファイル全体の文字起こし、発言者の分類、各発言の発話開始時刻の取得を行う。

②性別の特定

取得した発話開始時刻をもとに、録音ファイルを発言ごとに分割する。分割した各発言者の音声ファイルの冒頭5秒間のみを切り抜き、スペクトログラムに変換する。学習させた畳み込みニューラルネットワーク(CNN)に音声ファイルから変換したスペクトログラムを入力し、性別を特定する。

③ストーリー文章と助言の生成

全体の録音ファイルから文字起こしした内容をもとに、それぞれの発言者の発言を要約した文章であるストーリー文章と各発言者に対する客観的視点からの助言をテキストファイルとして生成する。ストーリー文章には性別情報も含まれる。

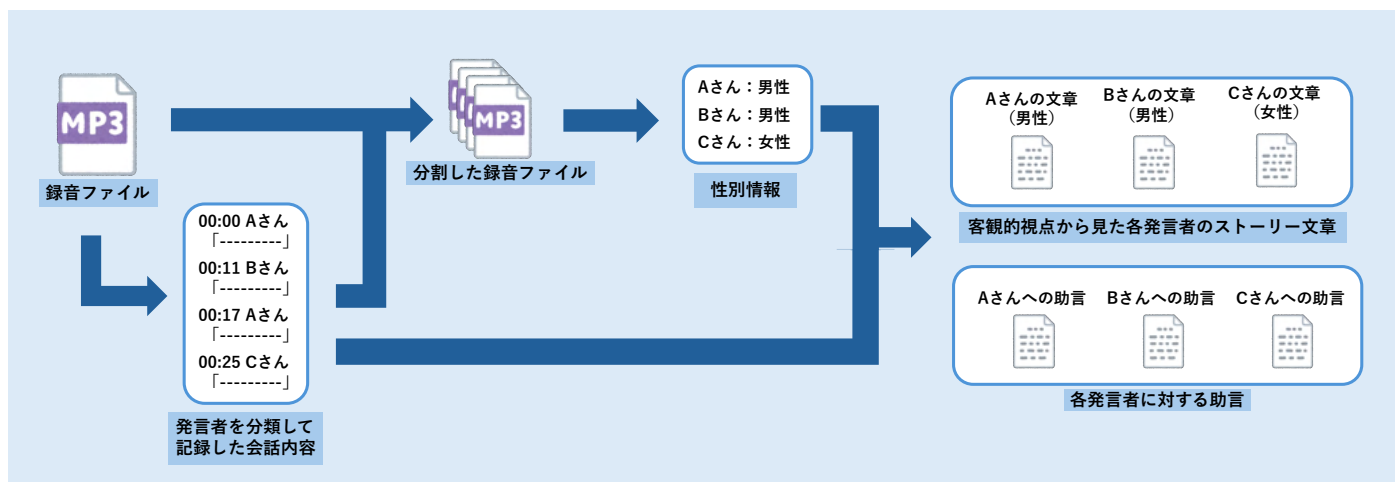


図3 文字起こしの流れ

● 漫画生成

①1コマ目の画像生成

生成されたストーリー文章をもとに、画像生成AIであるStable Diffusionを用いて漫画の1コマ目の画像を生成する。

②漫画の続き生成と助言の提示

出力された1コマ目の画像とストーリー文章をマルチモーダルLLMであるSEED-Story[6]に入力する。SEED-Storyはストーリー文章と物語の開始部分となる画像を入力として物語に沿った画像を出力する。①で生成した漫画の1コマ目の画像とストーリー文章から、画像に一貫性があり、ストーリー文章に忠実な漫画を生成する。漫画は各発言者に対してそれぞれ生成されるため、事前に登録したユーザ情報から、音声情報をスペクトログラムに変換し、CNNを用いて会話における発言者の音声と照合を行い、合致した漫画をユーザに提示する。

会話に参加した田中さんに対して生成される出力の例を図4に示す。田中さんは責任感が強く、部署のリーダーとして活躍する男性社員である。客観的視点から見た田中さんの様子と会話内容、そして発言内容に対する助言を提示する。

漫画では、現実の自身や他者をキャラクターとして表現するため、自然と自身とキャラクターの間に距離感が生まれる。この距離感により自己を客観的に捉えることができる。キャラクター化された自身を観察することで、自分の行動や発言を第三者として捉えることができ、文字だけの議事録や文章よりも客観的視点を得やすくなる。



図4 田中さんが受け取る漫画の生成例

田中さんへの助言

- ・田中さんはいるも責任感をもって仕事に取り組んでいて素晴らしいです！チームメンバーの意見も参考に、より良い戦略を検討してみتهいかがでしょうか？
- ・市場調査の結果を参考に、ターゲット層をより広げること、さらなる成功の可能性を広げることができるかもしれません。

展望

試作中のシステムでは利用シーン、機能が限定されている。また漫画の生成に用いる文章は録音内容を単純に文字起こしした内容から生成しているため、生成される漫画のクオリティに限界がある。これらの課題に対し、以下のようなアプローチでシステムの改善を図る。

○汎用性

様々な会話のシチュエーションにおいても本システムを使用することで手軽に自身の振り返りができるようにする。手軽に持ち運び可能なウェアラブル端末から会話内容を録音することで日常会話における振り返りが可能となる。

○漫画のクオリティ向上

ある特定の人物に向けた漫画を作成する場合に、その人物の発言に対して他者がどのように感じるのかを感情認識を用いて漫画の内容に組み込むようにする(Aの発言に対しBが顔をしかめている様子を表現する等)。これにより、本システムでは「助言」として漫画の他に付加していた内容を、漫画内容に組み込むことが可能である。

○機能の発展

漫画を作成する際に、特定の他者からの視点の漫画を作成する機能を追加する。これにより他者視点から見た自分の発言を考察することができ、今後の行動改善に役立てることが可能である。

参考文献

- [1]小磯花絵, 渡部涼子, 土屋智行, 横森大輔, 相澤正夫, 伝康晴, 一日の会話行動に関する調査報告, 国立国語研究所「日常会話コーパス」プロジェクト報告書1 (2017)
<https://www2.ninjal.ac.jp/conversation/report/report01.pdf>
- [2] The Crystal Approach to Developing Software, <https://slideplayer.com/slide/10167232/>
- [3] YWT (やったこと・わかったこと・次にやること), <https://www.jmac.co.jp/glossary/n-z/ywt.html>
- [4] 西川大志, 松永美希, 古谷嘉一郎, 反すうが自動思考と抑うつに与える影響, 心理学研究, Vol. 84, No. 5, pp.451-457 (2013)
<https://doi.org/10.4992/jjpsy.84.451>
- [5] Severance, C., Hanss, T., and Hardin, J.: IMS Learning Tools Interoperability: Enabling a Mashup Approach to Teaching and Learning Tools, Technology, Instruction, Cognition & Learning, Vol. 7, No. 3-4, pp. 249-262 (2010)
<https://www.dr-chuck.com/dr-chuck/papers/2009/ticl-01.pdf>
- [6] Shuai Yang, Yuying Ge, Yang Li, Yukang Chen, Yixiao Ge, Ying Shan, Yingcong Chen, SEED-Story: Multimodal Long Story Generation with Large Language Model, (2024)
<https://arxiv.org/abs/2407.08683>