報告

オンライン時代のコミュニケーション支援AIの設計

A design of communication support AI in the online era

橋本 昌嗣 HASHIMOTO Masatsugu
デジタルハリウッド大学大学院 客員教授 / 株式会社エヌジーシー 代表取締役
Digital Hollywood University, Graduate School, Visiting Professor / NGC CEO

堀 健一郎 HORI Kenichiro
株式会社飯綱エンタープライズ 取締役
Iizuna Enterprise,Inc Managing Director

小田 幸弘 ODA Yukihiro
株式会社エヌジーシー 開発部 部長
NGC Development Division, Manager

出雲 正尚 IZUMO Masanao
株式会社エヌジーシー 開発部
NGC Development Division

北岡 泰典 KITAOKA Taiten
株式会社オフィス北岡 代表取締役
Office Kitaoka, Inc. CEO

大塚 寛 OTSUKA Hiroshi
PST株式会社 代表取締役 / デジタルハリウッド大学大学院 ゲスト講師
PST, Inc. CEO / Digital Hollywood University, Graduate School, Guest Lecturer

光吉 俊二 MITSUYOSHI Shunji
東京大学大学院 工学系研究科 特任准教授
School of Engineering, The University of Tokyo, Associate Professor

コロナ禍の中、急速にコミュニケーションのオンライン化が進んだ。オンラインは対面よりも相手の感情がわかりづらい。相手の感情を把握するために、NLPと呼ばれるコーチングの手法をAI化することを堀が考案した。NLPは、カルフォルニア大学サンタクルーズ校で生み出された手法で、コミュニケーションが卓越した人たちの手法を体系化したものである。われわれは、この手法をAIで支援することを目的とし、NLPのエキスパートである北岡の助言をもとに、橋本がAI化の設計を行った。1つ目は、人が発する言葉の中に、「視覚」「聴覚」「その他の感覚」に関する単語が含まれるかを形態素解析によって検出し、どの感覚が強いかを検出することを可能とした。2つ目は、ロボットのペッパーに搭載されている光吉が開発した「感情エンジン」、および大塚が開発した「MIMOSYS」により、声からの喜怒哀楽と元気圧の検出を実現した。それに加え、3つ目として、顔の表情からの喜怒哀楽、笑顔の検出を、小田、出雲にて実装した。これにより、声および表情より感情を把握することを実現した。NLPの手法には、コミュニケーションがシンクロしていることを確認する手法や、相手の意識を変えることまで含まれている。また、微表情の解析も重要だと考えている。これらが将来の研究課題である。

キーワード：コーチング、NLP、AI、感情認識、表情認識

1. はじめに

1.1 背景

コロナ禍で、コミュニケーションの多くがオンラインにシフトした。社内会議、社外とのミーティング、商談、授業、学会発表、展示会、面接、同窓会等全てオンライン化した。その中で、オンライン会議は、対面よりわかりづらいと聞く。オンライン面接等を支援するサービスも開始されている[1]。

1.2 NLPとは

日本のメンタルヘルスは欧米に比べて遅れている。米国では、1970年代、ベトナム戦争の帰還兵の多くがPTSDに悩まされ、その短期的治療法として、Neuro-Linguistic Programming（NLP）と呼ばれる、コーチングのメソッドが利用されてきた。NLPとは、カリフォルニア大学サンタクルーズ校のジョン・グリンダーとリチャード・バンドラーがセラピストの認知・行動パターンをモデル化したものである。モデル化されたNLPの技法は、体系的にまとめられ、極めて科学的である。NLPのメソッドを使用すると、例えば、コーチがクライアントから瞬時に信頼を得て、クライアントの無意識的な能力を引き出すことができる。その一方で、コミュニケーションの上級者の手法を体系化し、他の人々にその知識の活用支援を実現した。

2. コミュニケーション支援AIの設計

パターン化されたNLPの技法について、堀は、ルールの定義やAIによってコミュニケーションの支援に使えると考えた。

2.1 コミュニケーション支援AIのアーキテクチャ

今回、開発した感情認識エンジンの概要を図1に示す。図中のピンクの部分が今回独自開発した部分である。

まず、1つ目の「VAK診断」（視覚・聴覚・その他の感覚のどの感覚が強いかの検出）機能であるが、人は、視覚、聴覚、その他の感覚のどの感覚が強いかによって分類することが可能である。会話の内容を文字起こしし、使用する単語が視覚・聴覚・その他の感覚のどれに関連するものかを判定し、その人がどの感覚が強いかの判別を実現する。

2つ目の「自信解析」機能は、声による感情の分析とAIによる表情の分析の組み合わせにより、自信の分析を行い、自信があるかどうかを検出する。顔の表情から取得する情報は、顔468点および目76点x2つの箇所のメッシュ情報から、次の情報を用いて推測する。目に関しては、目線の左右･上下、目の開き具合を解析し、顔の向きに関しては、左右、上下、傾きを解析、顔全体に関しては前後の移動を解析、口に関しては、広がり具合、開き具合の解析を行う。利用するカメラの性能、コンピュータの性能を合わせ、処理点数を増やし、精度を向上させていく。

©2021 Digital Hollywood University

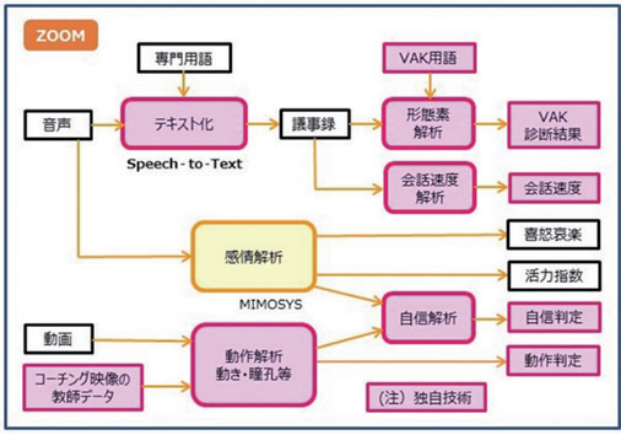

図1:感情認識エンジンの概要

３つ目の「信頼関係が構築できているかの検出」機能は、会話の際、NLPの日本での第一人者のアドバイスを受け、コーチングビデオからタグが付加された情報をAIに学習させ、動作がシンクロしているかなどから、お互いの信頼関係が構築できているかを検出する。

４つ目の「信頼を構築し、次のアクションの提示」機能は、会話の際、信頼関係が構築できた上で、次にどのようなアクションを起こせばよいかをAIが支援する。「VAK診断」「会話速度」「喜怒哀楽」「活力指数」「自信判定」「動作判定」情報を使い、次のアクションを提示する。

これらの4つの機能により、感情を読み取り、従来の自然言語処理では対応できなかった人間らしい会話を実現する。これを用いて、将来、次世代のアバターの実装を考えている。

これらの情報を統合して、図２に示すような形で、情報を提示することを目指す。

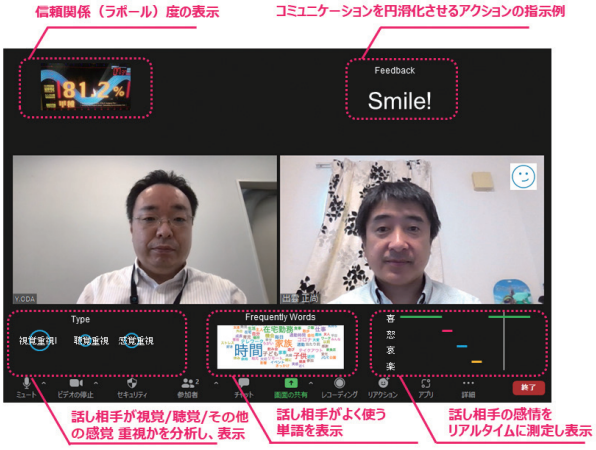

図2：感情認識エンジンが目指す提示情報のイメージ

提示できる情報としては、相手が視覚、聴覚、その他の感覚のどれが強いか、相手がどんな単語をよく使うか、 光吉が開発した感情エンジンを使って相手の感情の検出、信頼関係がどのくらい構築できたかの表示、その上で、どのようなアクションを取ればよいかをレコメンドする機能を考えている。

3. 感情認識エンジンの実装

3.1 語彙からの強い感覚の検出

NLPの１手法として、実世界を捉える際、視覚、聴覚、その他の感覚のどの感覚が鋭いかで、その捉え方が人によって異なると言われている。どの感覚が強いかは、どのような単語を使って話しているかを分析することによってわかる。例えば、視覚の感覚の強い人は、「その話はよく見える」、聴覚の強い人は、「その話はよく聞く」、その他の感覚の強い人は「その話は腹落ちする」といったような表現をする。視覚、聴覚、その他の感覚に関わる単語は表１に示すようなものがある。

表1：視覚、聴覚、その他の感覚に関わる単語

	それぞれに関わる単語
視覚	イメージ、イラスト、かすむ、くっきり、ご覧ください、ハイライト、はっきり、パノラマ、ビジュアル、ビジョン、フォーカス、プレビュー、まぼろし、暗い、闇、映え、垣間見る、外観、観察、見える、見かけ、見つめる、見通し、見張る、光景、公演、視覚、錯覚、私の視点、照明、色、鮮明、鳥観図、背景、反映、明確な、明示するなど
聴覚	アクセント、アコースティック、おしゃべり、きしむ、コメント、ざわつく、スピーチ、ノイズ、バズる、ピッチ、叫ぶ、響く、語る、呼ぶ、告げる、雑音、耳、声、静か、奏でる、前代未聞、宣言、しゃべる、聞く、聴く、相談、閑か、調べ、同調、波長、調子、調和、反響、悲鳴、無言、鳴く、余韻、和音、唸る、不協和、不調和など
その他の感覚	いじる、ウェット、おいしい、かぐわしい、キャッチ、くすぐる、こする、さする、しがみつく、ジャンプ、ショック、しょっぱい、ストレス、スムーズ、ストップ、すり抜ける、スルー、ぞくぞくする、ぞっと、タイト、タッチ、タイプ、ため息をつく、つかむ、つなぐ、ドキドキする、なでる、におい、ねじる、ハード、バランス、フィーリング、プレッシャー、まっすぐ、よろめく、リラックス、握る、圧力、威圧、引く、鋭い、押す、温かい、温和、回す、壊す、快適、気分、強い、苦い、軽い、結び付ける、固い、絞る、支える、持つ、指すなど

これらは、科学的な手法で分析可能である。まず、会話の内容をAIを用いて、Speech-to-Textにより文字に起こす。その文中において、視覚、聴覚、その他の感覚に関する単語がどのくらいあるのかを、形態素解析を用いて分析すると、その人がどの感覚が強いかが分析可能である。コーチングのメソッドでは、相手が使う感覚に関する単語使うほど、早期に信頼関係が築けることがわかっている。また、同じ感覚が強い人がコーチングをすると、相性がよいともされている。

3.2 声からの喜怒哀楽の検出

ロボットで会話を成立させるためには、自然言語以外の声のトーンや顔の表情から感情を読み取ることが必要である。光吉は、脳と密接な関係にある声帯から発生される声のトーンより、喜怒哀楽の感情を分析する感情エンジンを発明した。この感情エンジンはSoftbank社のロボット「ペッパー」にも搭載され、人間の声のトーンを読み、軽妙に会話をつなげることができる。また、東日本大震災で遺体処理をした自衛隊員の方々の中には鬱になる方も多く、鬱の判定において、医師の診断、血液検査の結果、感情エンジンの結果が一致したことも示した。

3.3 音声からの元気圧の検出

大塚は、感情エンジンをさらに発展させ、本人でもなかなかわかりづらい、常に変化している心の状態を声から検出する「元気圧」の測定を実現した。元気圧とは短期的な心の状態のことである。大塚は、声帯の不随意反応に着目し、声の周波数の変動パターン等から心の状態の分析を実現した。例えば、人前での発表で緊張し、声が上ずるような声帯の緊張は自分では操作することはできない。これが声帯の不随意反応である。定型的に元気圧を測定すると、自然災害が発生した地域の方々の元気圧が下がっていることなどがわかった。

©2021 Digital Hollywood University

3.4 顔の表情からの感情と笑顔の検出

本実装においては、小田と出雲がオープンソースのBlazeFaceを用い、顔がどこにあるのかを見つけ出し、face-api.js(Facial Expression)を用いて表情の推定を図３に示すとおり、行った[2]。BlazeFaceとface-api.jsとの連携は、本研究独自のものである。また、iPhoneで動作させるときには、iPhone内のAIチップにより高速化する実装も行った。

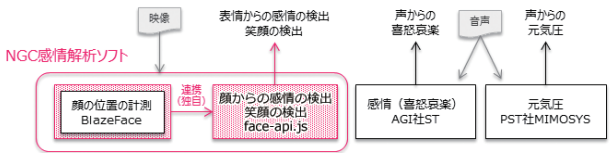

図3：顔の表情からの感情と笑顔の検出ソフトウェアの構成

(1) 顔の検知

Facial Landmarks Detectionを利用し、視線の向き、顔の向き、口の開き具合を測定し、数値化してグラフに表示することを図４のとおり実現した。

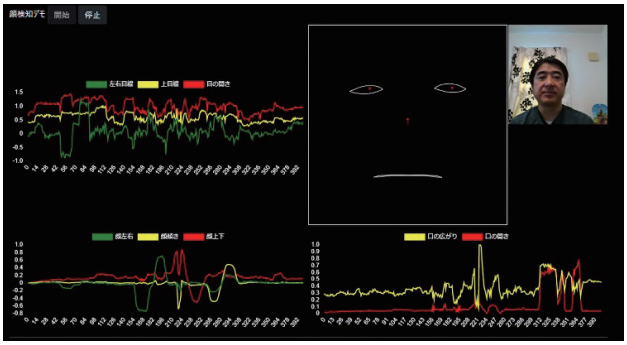

図4：実装した顔の表情の測定画面

(2) 笑顔の検知

笑顔の検知に関しても、Facial Landmarks Detectionを利用し、検出を実現した。笑顔を検出した場合、図５のとおり、エフェクトを付加する機能も実装した。FER+のデータセットを利用し、笑顔の評価を実施したところ、78%の精度を得た。

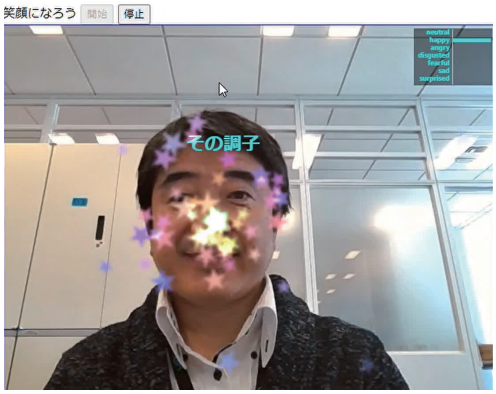

図5. 笑顔の検出

3.5 声および表情により感情を取得できる一覧

今回、実装し、感情を取得できた一覧を表２に示す。

表2：感情を取得できた一覧

ソフトウェア名	取得できる情報
VAK診断	発する言葉を形態素解析し、視覚、聴覚、その他の感覚に関する単語を使用しているかの検出
感情エンジンおよびMIMOSYS	声より感情解析：楽・怒・悲・平坦 元気圧・鬱・アルツハイマーの可能性
BlazeFace	映像中の顔の位置検知
face-api.js (Facial Expression)	表情解析：笑顔・怒り・軽蔑・恐怖・悲しみ・驚き・平坦
組み合わせでできそうなこと	アプリ上で声と表情からとれる感情をそれぞれユーザーに提示。感情と表情の関連やギャップが視覚的にわかる

4. 今後の課題

感情を認識する上で、「微表情」の認識が重要であると考えている。「微表情」とは、抑制された「真の感情」が顔に表れてフラッシュのように一瞬で消え去る現象のことである。その多くは0.2秒以内の出来事で、通常の会話では80～90%が見落とされていると言われている。微表情には表３に示すようなものがあると言われている[4]。

北岡のメソッドを補完する上で、堀は、今後、その微表情も解析したいと考えている。そのためには、Facial Action Coding System（顔面動作符号化システム）も実装したいと考えている。

表3：微表情の種類[4]

	微表情			
感情	眉	目	口	鼻
幸福	-	笑いジワができ、頬が上がる	口角が上がる	-
軽蔑	-	-	片方の口角が上がる	-
嫌悪	-	-	上唇が上がる	鼻の周りにシワがよる
怒り	両眉が中央によりながら下がる	目を見開き、下まぶたが力む	口に力が入る	
悲しみ	眉の内側が上がり、ハの字になる	-	口角が下がり、下唇が上がる	-
恐怖	両眉が上がり、中央による。額中央に波上のシワができる	目を見開き、下まぶたが力む	口角が横に引かれる	-
驚き	両眉が下がり、額全体に水平のシワができる	見開く	口を開ける	-

5. 結語

本研究では、発する語彙からの視覚・聴覚・その他の感覚のうちでどの機能が強いかの検知、声からの感情解析、表情からの感情および笑顔の検出を実現した。実装を行ったが、多くの被験者を使っての評価はまだできていない。

光吉、大塚による研究より、声帯は脳に近く、声には無意識な感情が出やすいことがわかっている。それに対し、顔の表情については、人によって表情に出る。あるいは、出ないといった言葉があるように、顔の表情には感情が出ないこともある。これらを踏まえ、人の感情をいかに捉え、活用していくかが今後の課題である。

また、コーチングの手法の１つであるNLPでは、人の視線、表情、しぐさなどから相手の心理状態を察知することができる。それにより、短期間で信頼関係を構築し、内面に向き悩んでいる相手の意識を、現実世界に向けさせ、直視させることによって、PTSD等を治療してきた。優秀なコーチングのトレーナーは、相手の意識を変えるタイミングをチョイスポイントとし、検出することができる。このような手法の実装が今後の課題である。

参考文献

[1] https://zenkigen.co.jp/（2021年9月20日）
[2] https://github.com/justadudewhohacks/face-api.js/
(2021年9月20日)
[3] https://arxiv.org/abs/1907.05047　(2021年9月20日)
[4] https://careersupli.jp/career/honne/　(2021年9月20日)