



人の思考力を理解・再現・拡張するための思考処理技術

NTT人間情報研究所（人間研）思考処理研究プロジェクト（思考P）では、「情報理解」「情報処理」「反応・行動」で構成される人の思考力について理解を深め、それらを計算機上で再現すると同時に、人の思考力を持続成長可能とする技術創出をめざしています。本稿では、この技術につながる「視覚的機械読解技術」「行動モデリング技術」「音声認識技術」「思考拡張型刺激デザイン技術」について紹介します。

にしだ	きょうすけ	くらしま	たけし
西田	京介	倉島	健
みやざき	のぼる	とだ	ひろゆき
宮崎	昇	戸田	浩之
にしおか	しゅういち		
西岡	秀一		

NTT人間情報研究所

はじめに

NTT人間情報研究所（人間研）思考処理研究プロジェクト（思考P）では、人の知覚から認知に至る「情報理解」、取得した情報を解釈し次の行動を起こすための「情報処理」、外界への働きかけである「反応・行動」、の理解を深め、それらを計算機上で再現すると同時に、人の思考力を持続成長可能とする技術創出をめざしています。この技術につながる4つの技術（文書を視覚的に理解する視覚的機械読解技術、行動から人の判断の仕組みを考える行動モデリング技術、人の内面を理解するための音声認識技術、人の思考力を引き出し拡張する思考拡張型刺激デザイン技術）について、紹介していきます。

文書を視覚的に理解する「視覚的機械読解技術」

人間研では情報検索や対話・質問応答サービスのさらなる発展をめざして、AI（人工知能）が自然言語（私たちが

日常的に用いる言葉）で書かれたテキストを読み、その意味を理解する「機械読解」の研究に取り組んできました^{(1)・(2)}。機械読解の研究は、一部の評価データにおいてはAIが人間の読解力を超えるなど大きく発展してきましたが、テキスト情報しか理解できないという限界がありました。一方で、私たちが普段扱っているPDF文書やプレゼンテーションスライドには、言語情報のみならず、文字の大きさや色、図や表、グラフ、レイアウトの情報などさまざまな視覚的要素が含まれています。視覚と言語の統合的な理解は、オフィス作業や日常生活を支援するAIの発展に向けて必要不可欠といえます。

そこで人間研では、文書画像に対する視覚的機械読解（図1）の実現をめざして、VisualMRCというデータセット⁽³⁾を構築して研究を進めています。このデータセットはWebページのスクリーンショットの文書画像に対する自由記述型の質問応答データであり、文書内の領域をタイトル・段落・リス

ト・画像・キャプションなど9クラスに分類してアノテーションした点が特徴的です。私たちはこのデータセットを用いて研究を進め、物体認識技術を適用して抽出した文書中の領域と、さらに文字認識技術を適用して抽出したトークンの位置・外観情報を追加入力として考慮可能な視覚的機械読解モデルを提案しました⁽³⁾。この技術は、人間の質問応答精度にはまだ及ばないものの、テキストのみを扱うモデルに比べて文書の視覚情報を理解することで質問応答の性能が向上することが確認されています。人間研では今後も視覚と言語の統合的な理解に向けてさまざまな研究に取り組んでいきます。

行動から人の判断の仕組みを考える行動モデリング技術

私たちは、行動経済学等の人文・社会的知見と、昨今のIoT（Internet of Things）の普及により得られるようになったさまざまな人のデータとの両面を活用しながら、不合理な判断を含む人の意思決定・行動判断のメカニ

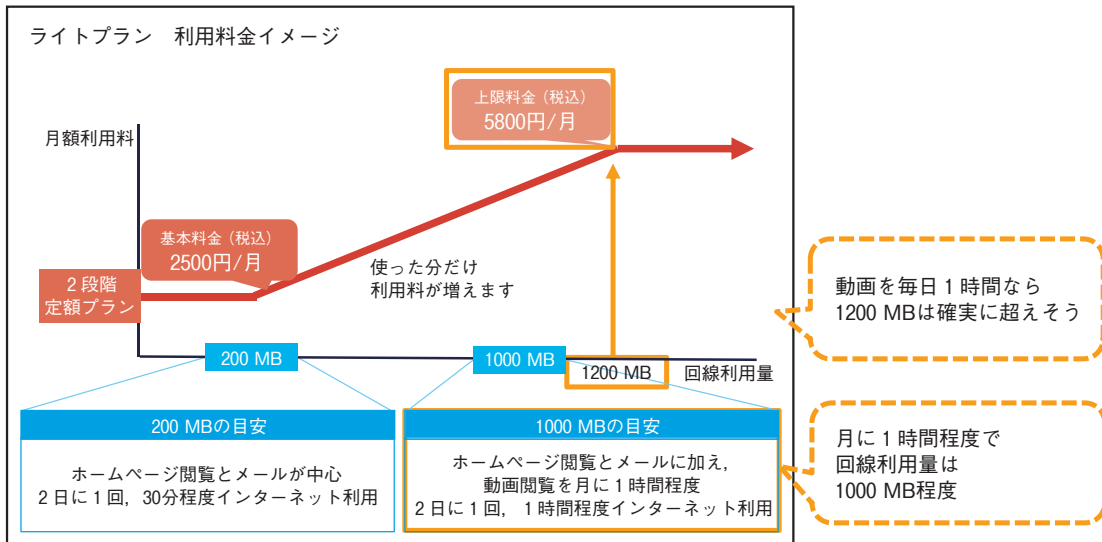
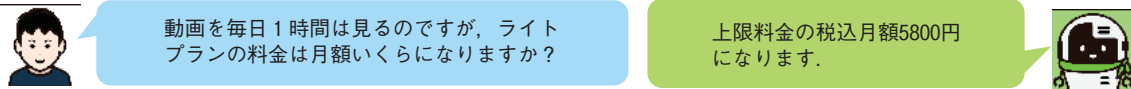


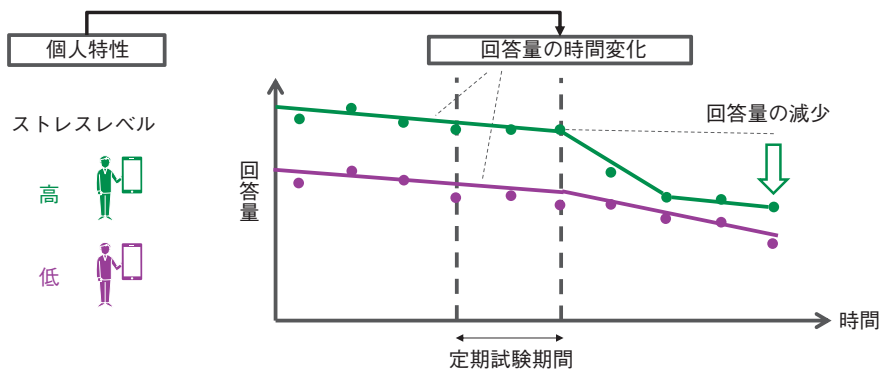
図1 視覚的機械読解

ズムを研究しています。さらに、得られた知見を骨組みとして人の意思決定や行動を再現する「行動モデル」を構築し、人に関する将来の予測やシミュレーションを可能とすることをめざしています。悪い未来が予測される場合には、より良い未来へと導く方策を、シミュレーション結果を使いながら探索することができるようになると考えます。

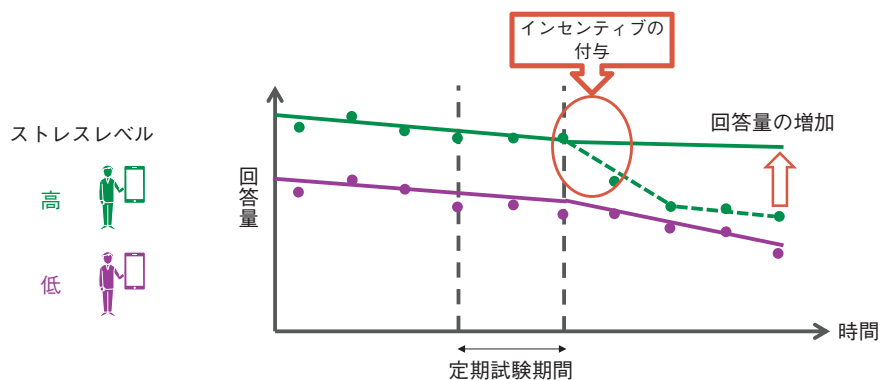
これまで、人の判断メカニズム解明の一環として、性格や価値観、身体的・心理的な状態、社会環境などの要因が個人の意思決定や行動に及ぼす影響について研究してきました。スマートフォンやタブレットなどのポータブルデバイスを通じて被験者に簡易な質問を送り、

そのときその場の状況、思考や感情、そして行動について被験者に自発的に回答してもらう調査方法として、Ecological Momentary Assessment (EMA、別名で経験サンプリング法)があります。大学生を対象として実施された10週間にもわたるEMAアンケート⁽⁴⁾の結果を用いて、個人特性ごとのEMAアンケートへの回答傾向の異なりを「時間変化」という観点から詳細に分析しました⁽⁵⁾。例えば、EMAアンケート調査開始前の事前調査で日常的にストレスレベルが高いと判定された被験者は、図2(a)に示すように、調査期間中もデバイス・機械に対して積極的に自己開示する傾向がありました。しかし、ストレスがかかるイベン

ト(例えば定期試験)の直後は回答に非協力的になり、急激に回答量が落ち込む傾向があることが私たちの分析により分かってきました。また、事前調査で誠実性が高いと判定された被験者は、EMAアンケート調査が始まった直後は協力的で回答量も多い一方、時間の経過に伴い回答量が大きく減少する傾向があることも分かりました。こういった人間行動(ここではアンケートに自発的に回答する行為)の時間変化に対する理解が深まることにより、将来どうなるのか、といった人の行動や結果を予測することができ、未来をより良い方向に変える方策の決定がしやすくなると考えています。EMAアンケートは、その人の状態を常に把握



(a) 個人特性（ストレスレベル）や社会環境（定期試験）と回答量の時間変化の分析



(b) 回答量が減少するタイミングでインセンティブを与え回答量の増加を促す

図2 行動データ分析結果に基づく方策の探索

人の内面を理解するための音声認識技術

スマートフォンやAIスピーカに話しかけて機器を操作する音声アシスタントの登場により、音声認識技術が世の中に急激に普及しました。このような、短い言葉のやり取りによる音声対話でコンピュータを操作するための音声認識技術の実用化は、古くは1980年代の自動音声応答装置や1990年代のカーナビゲーションへの導入などに始まり、近年では深層学習技術の導入による音声認識精度の大幅な向上によって、音声アシスタントの急激な普及につながりました。

一方、音声は人どうしのコミュニケーションにおける重要な情報伝達手段でもあります。人どうしの自然なコミュニケーションにおいては、音声アシスタントへ呼びかける短文音声とは異なり、比較的長い長文の音声が含まれ、より複雑な言語表現が現れることから音声認識にとっては対応が難しくなります。さらに、会話の相手が見知った相手になると、くだけた口調で断片的な発話が含まれるようになり、音としても、言語表現としても予測がさらに困難になります。一方で、このような音声を認識することで、コンタクトセンタに大量保存されている通話ログを対象とした会話内容の分析やリアルタイムの会話支援など、事業的に大きな価値のある用途に活用できるため、種々の深層学習を活用した音声認

するという観点から、時間や被験者に偏りなく、均一に回答数を確保できることが望ましいです。例えば、図2(b)に示すように、ストレスレベルの高い被験者に対して、回答量が落ち込むであろう試験終了後に備え、特別なインセンティブを与えて回答を促す、などの方策をとることが可能となります。

以上で示したのは、複雑な人間行動の一側面をとらえた分析に過ぎません。今後は、これまでに私たちが取り組んできた行動モデル⁽⁶⁾をベースに、

「不確実性に対する態度（リスク志向性）」や「待つことに対する態度（辛抱強さ）」など、行動経済学的人間特性も考慮して人の意思決定・行動モデルの高度化に取り組む予定です。なお、ここでいうモデルの高度化とは、モデルがより「人間らしい」判断をするようになることを意味します。「人間らしい」モデルだからこそ、悪い未来を含め、社会の未来の姿を予測し描きだせるのではないかと私たちは考えます。

識技術の検討が現在も活発に行われています。このように音声認識技術は、音声をテキストに変換する精度の向上とこれに伴う用途の拡大、扱う音声の複雑化が繰り返され、発展を遂げています(図3)。

一方、音声コミュニケーションを通じて伝達される情報には、言語情報(テキスト情報)だけでなく非言語的な情報(性別、年齢、感情、意図、態度など)も含まれています。私たちは、音声からテキスト情報を高精度に認識する取り組みとともに、非言語的な情報の認識・活用技術についても検討を進め、話者の属性(成人男性・成人女性・子ども)、感情(喜・怒・哀・平静)、疑問・非疑問を抽出できる技術を開発しています。また、コンタクトセンタ音声のようなオペレータと顧客との2名の話者からなる会話音声を対象として、顧客の怒り有無や満足度、オペ

レータ側の対応が顧客側に与える印象としての対応音声好感度を推定する技術の研究開発と実用化を進めています。

現在取り組んでいる、言語情報や非言語的な情報に対する認識技術は、より高度なサービスの実現に向けた、言葉に現れない人の心の状態を読み取る技術の第一歩です。私たちは普段の会話において、相手の感情や興味、好意や無関心などといった相手の心の内面にある情報を、声の調子や視線、表情、間の長さ、言葉遣いの変化などさまざまな手掛かりを使って推測することで、円滑なコミュニケーションにつなげています。このような人の内面情報を機械で認識することは、ユーザの気持ちに寄り添う音声対話エージェント、生徒1人ひとりの理解度や興味に応じた教育、問診対話データからの患者の体調や精神的ストレスの検出など、言葉に表れない人の心の機微を対

象にした新たなサービスにつながることを期待できます。さらに、NTTグループが進めるIOWN (Innovative Optical and Wireless Network) 構想の1つであるDTC (デジタルツインコンピューティング) が取り組むグランドチャレンジ「感性コミュニケーション」がねらう、経験や感性などの個々人の特性の違いを超え、心の中のとらえ方や感じ方を直接的に理解し合える新たなコミュニケーションの実現に貢献することが期待されます。

人の思考力を引き出し拡張する 思考拡張型刺激デザイン技術

石鹸のにおいをかぐと掃除がしたくなる⁽⁷⁾、商品パッケージの色を変えると、同じ中身でも異なる商品に感じる⁽⁸⁾ように、人の思考や行動は、言語や数値等のように意識的に解釈された情報の影響を受けるだけでなく、色や香り、話し方の雰囲気等のように、無意識的に感じる感覚や知覚の影響も受け、変化することが知られています。私たちのグループでは、「人が知覚する刺激」と「人の思考や行動」との関係性にかかわる知見を蓄積し、その関係性に基づいて、人の思考力をより引き出し拡張していくための知覚刺激生成・制御の研究に取り組んでいます。

これまでの取り組みの1つとして、発話者の話し方(声の高さ、話す速度、抑揚の大きさ)が人の心理的な状態や行動に及ぼす影響について調査した研

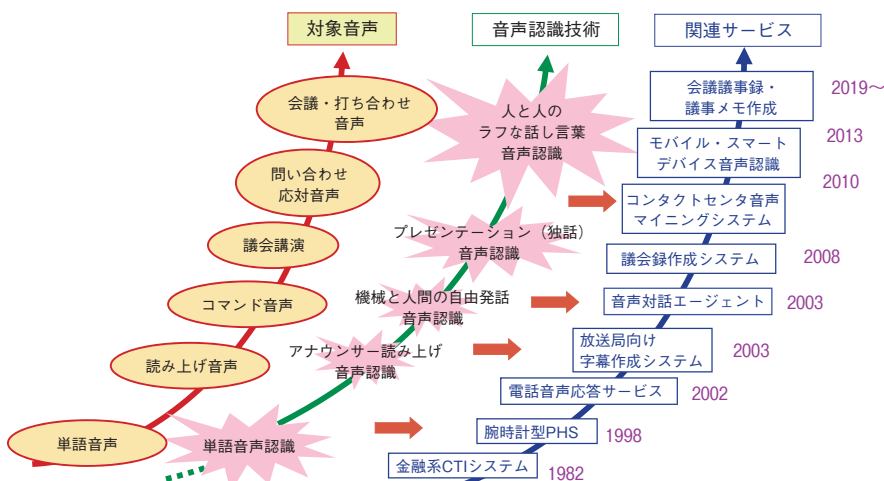


図3 NTTにおける音声認識技術の取り組み

究⁽⁹⁾があります。この研究では、商品を宣伝する音声と、人の消費行動との関係を対象として、大規模な主観評価実験を実施し、購買行動につながる「購買意欲」と「話し方」との関係性を、感情を媒介として表現する消費者行動モデル⁽¹⁰⁾（図4）を用いて分析し

ています。

■実験手順

クラウドソーシングにて、日本語を母語とする男女202名に、話し方を変えた電化製品に対する「宣伝文」を聴取してもらい、それを聞いた際に感じた「感情」と「購買意欲」について

評価項目に回答してもらいました。そして、得られた話し方（音声特徴）に関するパラメータ（声の高さ、話す速度、抑揚の大きさ）と、評価項目に対する回答を基に、音声特徴と感情、購買意欲の関係を分析しています。宣伝文は、プロの女性話者が読み上げ調で発話した音声を元に、表1に示すパラメータを調整した音声を利用し、評価項目は表2に示したものを利用しました。

■実験結果

消費者行動モデルに基づく3層モデルを利用して、音声特徴と、感情、購買意欲の関係性を分析した結果を図5に示します。結果を見ると先行研究⁽¹¹⁾でも示されているとおり、感情のうち、快感情と覚醒が購買意欲に影響を及ぼしていることが分かります。また、感情の中でも快感情の影響がもっとも強いことが分かります。また、声が高く、話速が早く、抑揚が大きいことが、感情の評価値が高くなることにつながり、その中でも話速の影響がもっとも大きいことが分かります。

これらの結果をまとめると、広告文の読み上げによる音声刺激においても「外部刺激⇒感情⇒購買意欲」のモデルが成立すること、話速がもっとも快感情に影響し、快感情がもっとも強く購買意欲に影響することが分かりました。

以上で示した取り組みは、複雑で多様な人の思考プロセスからすると、そのごく一部を取り上げた分析に過ぎま

消費者行動を 外部刺激⇒感情⇒行動 の3ステップで表すモデル

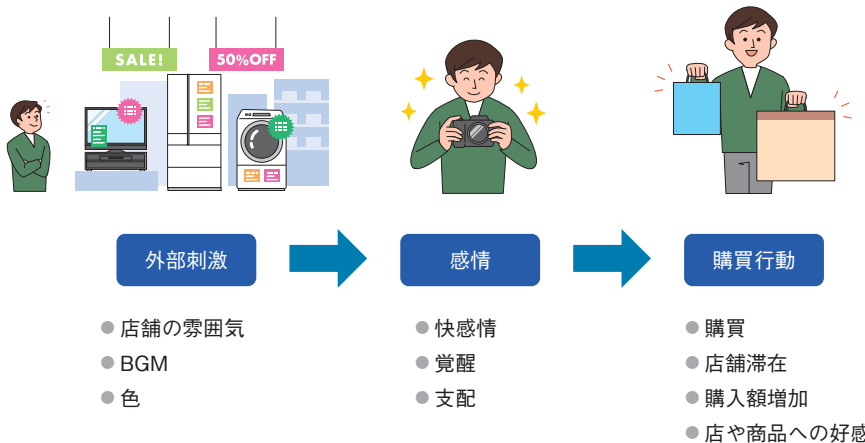


図4 消費者行動モデル

表1 音声のパラメータ

音声のパラメータ		設定値
声の高さ	平均F0 [Hz]	0.9439倍, 1倍, 1.059倍
話す速度	話速 [mora/s]	0.890倍, 1倍, 1.1225倍
抑揚の大きさ	F0分散 [Hz]	0.6667倍, 1倍, 1.5倍

表2 評価項目

評価項目		設定値
感情	快感情	不快⇔快 (7段階)
	覚醒	睡眠⇔覚醒 (7段階)
	支配	服従⇔支配 (7段階)
購買意欲		非常に購入したい⇔全く購入したくない (7段階)

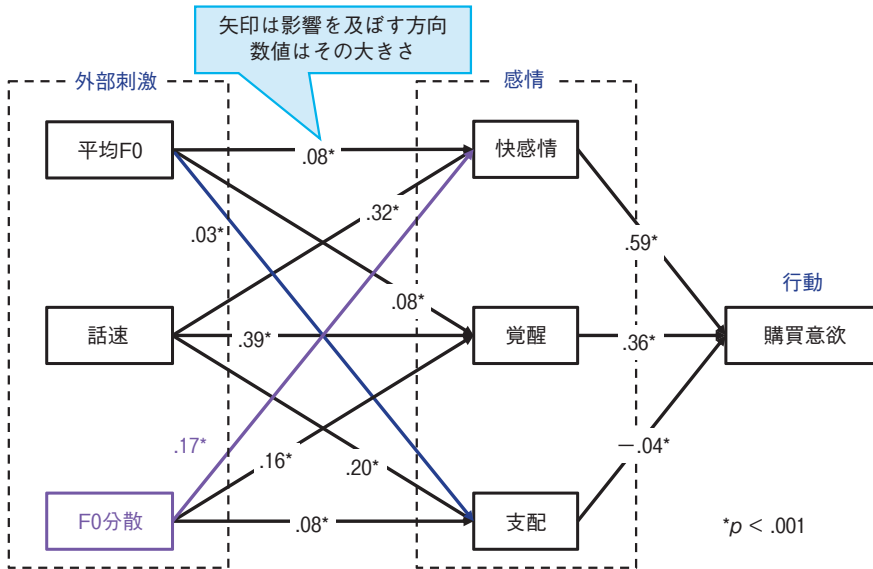


図5 音声特徴、感情、購買意欲の関係分析結果

せん。今後は、聴覚刺激においても聴取者の属性との関係等、より幅広い要素について検討を進めるとともに、視覚や嗅覚など聴覚以外の刺激についても取り組む予定です。また、刺激により影響を与える人の思考現象も、直感的な判断から、論理的思考や人の価値観への影響等より深い段階への影響について検討を進める予定です。これらの研究を通じて、人の思考力を引き出すために何が必要かを明らかにし、人の可能性を高め、より良い社会の実現に貢献したいと考えます。

おわりに

今後は、前述した4つの技術を磨き上げることで、「情報理解」「情報処理」「反応・行動」で構成される人の思考力を理解・再現・拡張していく方針

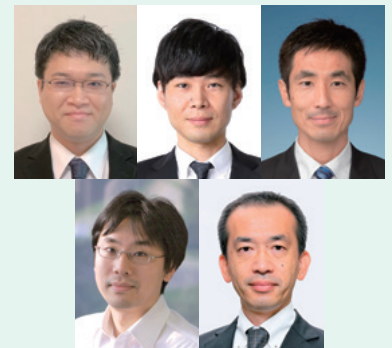
です。

参考文献

- (1) K. Nishida, I. Saito, K. Nishida, K. Shinoda, A. Otsuka, H. Asano, and J. Tomita: "Multi-style Generative Reading Comprehension," ACL 2019, pp. 2273-2284, 2019.
- (2) K. Nishida, K. Nishida, M. Nagata, A. Otsuka, I. Saito, H. Asano, and J. Tomita: "Answering while Summarizing: Multi-task Learning for Multi-hop QA with Evidence Extraction," ACL 2019, pp. 2335-2345, 2019.
- (3) R. Tanaka, K. Nishida, and S. Yoshida: "VisualMRC: Machine Reading Comprehension on Document Images," AAAI 2021, pp. 13878-13888, 2021.
- (4) R. Wang, F. Chen, Z. Chen, T. Li, G. Harari, S. Tignor, X. Zhou, D. Ben-Zeev, and A. T. Campbell: "StudentLife: assessing mental health, academic performance and behavioral trends of college students using smartphones," Proc. of UbiComp 2014, Seattle, U. S. A., Sept. 2014.
- (5) T. Tominaga, S. Yamamoto, T. Kurashima, and H. Toda: "Effects of Personal Characteristics on Temporal Response Patterns in Ecological Momentary Assessments," Proc. of INTERACT 2021,

Bari, Italy, August 2021.

- (6) T. Kurashima, T. Althoff, and J. Leskovec: "Modeling Interdependent and Periodic Real-World Action Sequences," Proc. of WWW 2018, Lyon, France, April 2018.
- (7) K. Liljenquist, C. B. Zhong, and A. D. Galinsky: "The Smell of Virtue: Clean Scents Promote Reciprocity and Charity," Psychological Science, Vol. 21, No. 3, pp. 381-383, March 2010.
- (8) 槇久: "カラーデザインのための色彩学," オーム社, 2006.
- (9) M. Nagano, Y. Ijima, and S. Hiroya: "Impact of Emotional State on Estimation of Willingness to Buy from Advertising Speech," Proc. of INTERSPEECH 2021, 2021.
- (10) A. Mehrabian and J. A. Russell: "Approach to environmental psychology," The MIT Press, 1974.
- (11) R. Donovan and J. Rossiter: "Store atmosphere: an environmental psychology approach," Journal of retailing, Vol. 58, No. 1, pp. 34-57, 1982.



(上段左から) 西田 京介/ 倉島 健/ 宮崎 昇

(下段左から) 戸田 浩之/ 西岡 秀一

音声・言語処理等をベースに人の思考力を理解・再現・拡張することが可能な技術の研究開発を推進していきます。

◆問い合わせ先

NTT人間情報研究所
思考処理研究プロジェクト
E-mail shikou-inquiry-ml@hco.ntt.co.jp