特　集

デジタルとナチュラ
支えるコミュニケー

人工知能　クロスモーダル　シカクノモリ　雑談対話システム　CoCoNuTS

昨今，AI（人工知能）は特定の機能では人間の性能に迫るほどめざましく進歩しているが，まだ限定的で，一方人間は高度に複雑だが，それゆえにバイアスや錯覚に支配されるなど，不完全で誤りを犯す．本特集では，人間に迫るべくAI技術を研ぎ澄ませていくのと同時に，人間をさらに深く知ることで両者のギャップを埋め，人間に寄り添うAIを実現するためのコミュニケーション科学の取り組みを紹介する．

ルの共生・共創を
ション科学

人に迫り，人を究め，人に寄り添う ——デジタルとナチュラルの共生・共創に向けて

昨今，AI（人工知能）は特定の機能では人間の性能に迫るほどめざましく進歩していますが，まだ限定的です．一方人間は高度に複雑ですが，それゆえにバイアス（偏り）や錯覚に支配されるなど，不完全で誤りを犯します．本稿では，人間に迫るべくAI技術を研ぎ澄ませていくのと同時に，人間をさらに深く知ることで両者のギャップを埋め，人間に寄り添うAIを実現するためのコミュニケーション科学の取り組みを紹介します．

山田 武士（やまだ たけし）

NTTコミュニケーション科学基礎研究所 所長

はじめに

最近のAI（人工知能）技術の発展にはめざましいものがあります．もともとコンピュータは人間が処理できない大量のデータを一度に処理し，人間が苦手な処理を人間に代わって高速に処理するのが得意です．しかし特に深層学習の発展のおかげで，本来人間が得意で，なかなかコンピュータが追いつけなかった音声や画像の認識や自然言語処理などにおいても，人間の能力に迫り，場合によっては凌駕する性能を実現しつつあります．このようなメディア処理を中心に，今後さらにAIの進歩は加速すると期待されます．とはいえ脳の処理は複雑であり未解明の部分も多く残されています．AIの性能が複雑な人間の脳を超えるほどに進歩するのはまだ先といえます．

一方で人間は認知上のバイアス（偏り）にとらわれ間違いを犯したり，実際にはありもしない錯覚をリアルに感じてしまったりなど，複雑であるがゆえに一見すると不完全な存在でもあります．このように，限定された範囲で急速に発展を続けるコンピュータ（AI）と，複雑であるがゆえに不完全でもある人間とをつなぎ，両者のギャップを埋めることが「コミュニケーション科学」を研究所名に掲げるNTTコミュニケーション科学基礎研究所（CS研）の使命です（**図**）．これをふまえてCS研は人と人，あるいはコンピュータと人の間の「こころまで伝わる」コミュニケーションの実現をめざし，基礎理論の構築と革新技術の創出に取り組んでいます[(1)]．地道な基礎理論の構築の例としては，符号化効率の限界（シャノン限界）まで効率良くメッセージを送受信する符号化法の提案が挙げられます．こちらについては本特集記事『限界まで効率良くメッセージを送れます——シャノン限界を達成する通信路符号』で詳しく説明します[(2)]．今後さらに「こころまで伝わる」をめざすためには，メディア処理を中心とした人間の能力に迫る技術を追究するのはもちろんのこと，人間の機能，特性を解明し，人間のことをよく理解すること，そのうえで人間に寄り添う技術の実現をめざすことが一層重要であると考えています．

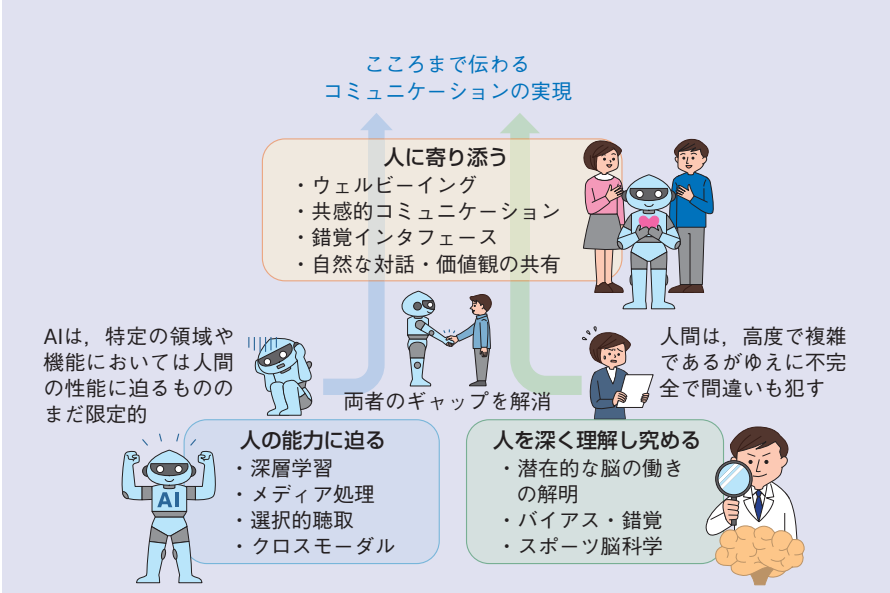


図 コミュニケーション科学のミッション

人間の能力に迫る技術

世の中にはまだまだ，人間は得意でも，コンピュータには苦手な処理が多数存在します．確かに機械翻訳の精度は飛躍的に向上し，大学入試の英語穴埋め問題をある程度正解できるようにはなりましたが[3]，文章の意味を深く理解したり，常識を身につけたり，というレベルにはまだ到達していません．

一方で，深層学習技術を駆使することで，画像認識や音声認識など，特定の面では人間の能力に迫ってきたことも事実です．例えば，会議やパーティでの歓談などにおいて，複数の人が同時に話したり，背景に音楽が流れていたりするとします．人間はこのような状況においても「聞きたい」人の声の特徴を選り分けて，話す内容を聞き取ることができます．これは人間の聴覚の優れた能力の１つで，選択的聴取と呼ばれます．選択的聴取はより広い概念である選択的注意の代表例です．従来，このような選択的聴取を，コンピュータは苦手でしたが，CS研では独自の深層学習技術により，人間同様，コンピュータが目的話者の声の特徴に基づき，その声だけを聞き取る技術を実現し，さらにそれを発展させています[4]．

これらのメディア処理技術が今後さらに進歩し，人間に近づくための鍵となるのがクロスモーダル処理です．クロスモーダル処理とは，「音声」「映像」「テキスト」など単一の「モダリティ」の垣根を越えた処理，という意味です．従来，これら「音声」「映像」「テキスト」などはそれぞれ解析手法も異なり，別々に研究されてきました．しかしここに来て，深層学習といういわば「共通言語」が整備されたおかげで，モダリティの垣根を越えた「認識」「生成」「変換」が可能になりつつあります．

一方，人間は常に複数の感覚（五感）を駆使して外界を知覚し，例えば，音声を聞いただけでその場の情景をある程度頭の中に思い浮かべることができるなど，このようなクロスモーダル処理を日常生活の中で当たり前のように行っています．また，目の見えない人が指先を使って点字を読むといった，障害などで損なわれた感覚の機能を残された感覚で代行する「感覚代行」もよく知られています．確かに人間なら，顔写真を見てその顔に合った声をある程度は想像できるかもしれませんが，そんなことがコンピュータに可能でしょうか？　CS研では実際に，これらのクロスモーダル処理をコンピュータで実現することに取り組んでいます．例えば「音から画像認識」するクロスメディア情景分析技術では，カメラでは死角になってしまうような個所の情報も音を使って「認識」できることをめざしています．CS研が取り組む最新のクロスモーダル処理技術については本特集記事『画像や音を見聞きするだけで賢くなるAI ――クロスモーダル情報処理の展開』で詳しく説明します[5]．

人間を深く理解し究める技術

このように特定の場面ではAIの能力は人間に近づき，凌駕しつつあります．しかし，AIの性能が複雑な人間の脳を超えるほどに進歩するのはまだ先でしょう．一方で人間は，「振り込め詐欺」にも簡単に騙されるなど，時として認知上のバイアスに支配されたり，錯覚にとらわれたりして，自分でも思いがけない誤りを犯します．CS研が運営するWebサイト「イリュージョンフォーラム」には自分の目や耳が信じられなくなるような，さまざまな錯覚の情報が掲載されています[6]．

クリストファー・チャブリスとダニエル・シモンズによる有名な実験[7]では，実験参加者は白シャツと黒シャツの合計６名の選手がバスケットボールをパスする映像を見せられ，白シャツチームの間でボールがパスされる回数を数えるように指示されます．このとき，映像の途中で９秒間かけて，舞台袖からゴリラが現れ，正面でカメラに向かって胸を叩き堂々と去っていくのですが，回数を数えるのに夢中の半数の実験参加者はそのことに気付きませんでした．このように人間はあることに注意を向けると，周囲で起こっている別のことに注意が向かなくなります．すなわち，人間の優れた特性である選択的注意は，裏を返せば選択的不注意であるわけです．しかもそうなっていることに本人は気が付きません．振り込め詐欺に騙されるのは高齢者だけとは限らないのです．

このように複雑であるがゆえに「バイアス」や「錯覚」にとらわれがちで不完全な人間と進歩しつつも今のところ限定的なAI，この両者のギャップを埋めて共生・共創していくためには，安易に「AIが人間の脳を超える」などと信じ込む前に，複雑な人間をもっと深く知ることが必要です．そのために，CS研では「視覚」「聴覚」「運動感覚」といった人間の基本的な感覚に関する「潜在的な脳の働き」の解明に注力しています．錯覚も「潜在的な脳の働き」の解明の重要な手掛かりです．

一口に脳の働きといっても人それぞれ多様です．CS研では，優れた運動能力を持つ一流アスリートに着目し，脳科学の視点から人間の「心・技・体」の関係の本質に迫る，スポーツ脳科学にも取り組んでいます．例えば，優れた打者がわずか0.1秒という短い時間で，いかに遅い球と早い球を見極めて球種に応じたタイミングで動いているか，その仕組みの解明などに挑んでいます．スポーツ脳科学は，ICTを駆使して主に体を鍛える従来のスポーツ科学や，パフォーマンスのみを評価するスポーツ分析手法とは一線を画した，野心的な取り組みです．

ちなみに，前述のクロスモーダル処理は脳内でもさまざまなレベルで行われています．例えば，通常の映像を見るとき，脳は映像中の「動き」「色」「形」（モダリティ）の情報を個別に処理し，後にそれらを統合します．したがって，これらの情報間に不整合があったとしても，統合する過程でそれは補正されます．この脳の処理の仕組みを利用してCS研で考案されたのが変幻灯®です[8]．変幻灯を体験するとき，ユーザは色や形は止まった対象から取得し，動きは投影されたモノクロの映像から取得します．色や形は止まっているので，動きと空間的に「ずれ」が生じます．しかし，辻褄が合ったようにものを見ようとする脳は，「動き」「色」「形」を統合する際に，その「ずれ」を補正します．そのため，変幻灯を体験する際には，ユーザは「動き」「色」「形」のずれに気付かずに，あたかも止まった対象の色や形が動いているように「錯覚」して感じるのです．

人間に寄り添う技術

スポーツ脳科学で得られる成果はスポーツに限らず，人間が普段の生活の中で心身の潜在能力を最大限に発揮する，すなわち，ウェルビーイングのための知見として活かすことができます．この人間のウェルビーイングという，一見，定性的でとらえどころのない課題を人間科学の立場から定量的に扱い，向上させるための設計指針の確立にもチャレンジしています．例えば，複数の人間が，場を一緒に共有することで生じる共感的コミュニケーションの効果測定などがその例です[9]．また，TVやスマートフォンなどのディスプレイ機器に日々囲まれて，ともすると目を酷使する現代人のために，汎用的なタブレット機器を用いてゲーム形式で日常的に目の状態をセルフチェックできる方法も提案しています．こちらは本特集記事『あなたの目の機能を気軽に楽しく測ります』で詳しく説明します[10]．

一方，錯覚は「潜在的な脳の働き」解明の手掛かりとして重要なのはもちろんのこと，人間とAIとのギャップを埋め，人間に寄り添うインタフェースやフィードバックのための鍵でもあります．CS研ではこれまで人間の錯覚を利用したインタフェースとして，引っ張られる錯覚を生じさせるデバイス「ぶるなび®」を考案しました．さらには，座っているのにあたかも歩いているような感覚の生成にも取り組んでいます．こちらは本特集記事『座っていても歩いているような疑似感覚の生成技術』で詳しく説明します[11]．また，前述の，印刷した絵や写真に光を当てるだけで動き出して見える「変幻灯」，3Dメガネを掛けると3D映像に，メガネを外すと鮮明な2D映像を楽しめる「Hidden Stereo」，印刷物などの2次元平面上の対象に対して影に見えるパターンを投影することで，その対象があたかも3次元的に浮き上がって見える光投影技術「浮像®（うくぞう）」[12]などを次々と生み出してきました．これからも，錯覚を利用した新たなインタフェースの提案と同時に，錯覚を通して物理的には生じ得ない体験を生み出すことによる，斬新な知覚表現の可能性も追究していきます．

ロボットやAIと人間との自然な対話を実現する，対話処理技術においては，重要なのは音声認識や自然言語処

理であって，一見，人間のバイアスや錯覚とは無関係にも思えます．しかしAIは人間のように文章の意味を深く理解したり，常識を身につけたりまではできないため，人間とAIの対話は現状ほぼ「一問一答式」に限られます．したがって，話していると少し前に言ったことと矛盾することを言うなどすぐボロが出て，対話は長続きしません．そこで，その限られた能力を効果的に活用しつつ人間のバイアスや錯覚を利用し，人間にとっていかに「賢く見せる」かが重要となります．CS研では2台のロボットでうまく役割分担をすることで，たとえ一問一答式であっても，自然な対話が長続きする対話処理を実現してきました．さらに一問一答式から脱却するために，ユーザの発話の多くがイベントに関する内容であることに着目し，イベント単位に構造化して把握する手法を提案しました．こちらは本特集記事『文脈を理解して話す雑談対話システム』で詳しく説明します(13)．こうすることで文脈理解度が向上するとともに，イベントにマッチするシステムの擬似経験も共有させることができ，その結果，ロボットに対する共感を誘発するなど，まさに人に寄り添う対話が可能になります．

おわりに

以上見てきたように，人間は高度で複雑であり，一方AIは特定の領域，機能においては人間の性能に迫るもののまだ限定的です．「人間を超える知能」はそう簡単には実現しないでしょう．しかし人間は複雑であるがゆえに不完全で，振り込め詐欺に騙されたり，因果関係を錯誤したり，選択的「不注意」とでも言うべきバイアスにとらわれ間違いを犯したりします．また，人間はありのままの物理量を見ているわけではないことが錯視例などからも分かります．以上のことからも，人間に迫るべくAI技術を研ぎ澄ませていくのと同時に，人間をさらに深く知ることによって，両者のギャップを埋め，デジタルとしてのAIが人間にナチュラルに寄り添い，両者が共生・共創する社会を実現することが重要であり，それが「こころまで伝わる」につながるCS研のミッションであると考えています．

■参考文献

(1) 山田：“新たな次元へとシフトする――さらに深化するコミュニケーション科学の取り組み，”NTT技術ジャーナル，Vol.30, No.9, pp.8-11, 2018.
(2) 村松：“限界まで効率良くメッセージを送れます――シャノン限界を達成する通進路符号，”NTT技術ジャーナル，Vol.31, No.9, pp.26-30, 2019.
(3) 東中・杉山・磯崎・菊井・堂坂・平・南：“「ロボットは東大に入れるか」における英語問題の回答手法，”NTT技術ジャーナル，Vol.27, No.4, pp.63-66, 2015.
(4) Delcroix・Zmolikova・木下・荒川・小川・中谷：“SpeakerBeam：聞きたい人の声に耳を傾けるコンピュータ――深層学習に基づく音声の選択的聴取，”NTT技術ジャーナル，Vol.30, No.9, pp.12-15, 2018.
(5) 柏野：“画像や音を見聞きするだけで賢くなるAI ――クロスモーダル情報処理の展開，”NTT技術ジャーナル，Vol.31, No.9, pp.10-13, 2019.
(6) http://www.kecl.ntt.co.jp/IllusionForum/
(7) http://www.theinvisiblegorilla.com/videos.html
(8) 河邉・吹上・澤山・西田：“変幻灯――止まっている対象を錯覚的に動かす光投影技術，”NTT技術ジャーナル，Vol.27, No.9, pp.87-90, 2015.
(9) 渡邊・大石・熊野・Hernández・佐藤・村田・麦谷：“ウェルビーイングを測る，知る，育む，”NTT技術ジャーナル，Vol.30, No.9, pp.29-32, 2018.
(10) 丸谷・細川・西田：“あなたの目の機能を気軽に楽しく測ります，”NTT技術ジャーナル，Vol.31, No.9, pp.14-17, 2019.
(11) 河邉：“座っていても歩いているような疑似感覚の生成技術，”NTT技術ジャーナル，Vol.31, No.9, pp.18-21, 2019.
(12) 河邉：“「浮像（うくぞう）」――影を利用して印刷物に見かけの奥行きを与える光投影技術，”NTT技術ジャーナル，Vol.30, No.9, pp.20-23, 2018.
(13) 成松・杉山・水上・有本・宮崎：“文脈を理解して話す雑談対話システム，”NTT技術ジャーナル，Vol.31, No.9, pp.22-25, 2019.

山田 武士

今後ますます技術の進歩のスピードが速くなり，競争が厳しくなる中で，CS研は，人に迫り，人を究め，人に寄り添う技術を中心に，これからも新たなチャレンジに大胆かつ粘り強く取り組んでいきます．

◆問い合わせ先
NTTコミュニケーション科学基礎研究所
企画部
TEL 0774-93-5020
FAX 0774-93-5026
E-mail cs-liaison-ml@hco.ntt.co.jp

画像や音を見聞きするだけで賢くなるAI ——クロスモーダル情報処理の展開

NTTコミュニケーション科学基礎研究所では，画像，音，テキストといった種類の異なるメディア情報にまたがる情報処理の研究を進めています．これをクロスモーダル情報処理と呼ぶことにします．クロスモーダル情報処理のポイントは，複数種類のメディアデータが対応付けられている共通の場所である「共通空間」をつくることです．これにより，これまでにはなかった新しい機能を実現できる可能性が示されつつあります．音から画像や説明文をつくるといった異種メディア間の新たな変換や，メディア情報に含まれる物事についての概念獲得などです．

かしの くにお
柏野 邦夫

NTTコミュニケーション科学基礎研究所

クロスモーダル情報処理とは

近年のAI（人工知能）の発展を支える立役者は深層学習の技術です．例えば，さまざまな物体を撮影した画像と「りんご」「みかん」といった物体の名前（クラスラベル）とを組（ペア）にしたデータを大量に用意して深層学習を行うと，画像中の物体が何であるかを高い精度で認識できるようになることが知られています．その優れた特性のためにさまざまな分野で研究や活用が進む深層学習ですが，私たちが特に着目している能力の１つは，異種のメディア情報（例えば，画像と音）の対応付けができることです．画像，音，テキストといった情報の種類のことをモダリティ（modality）と言いますので，異なるモダリティにまたがる情報の対応付けをクロスモーダル（cross-modal）情報処理と呼ぶことにします．このクロスモーダル情報処理とはどのようなもので，どんなメリットがあるのでしょうか．

新しい情報変換

⑴　音から画像をつくる

クロスモーダル情報処理のメリットの１つは，異種のメディア情報が対応付けられた共通の場所である「共通空間」を介することで，従来では考えられなかったような情報の変換が可能なことです（**図１**）．その１つとして，私たちの研究チームでは，音から画像を推定する課題に取り組んでいます．

私たち人間は，目を閉じていても周囲の音からその場の情景を思い浮かべることができます．そこで，マイクで拾った音からその場の情景を表す画像をつくってみようというわけです．例えば，室内に複数のマイクを設置し，数人の会話を数秒間録音します．４本のマイクを用いたとすると，それぞれのマイクでとらえた音の周波数成分の時間変化を表す「スペクトログラム」が４枚と，音の到来方向を表現した「角度スペクトル」の情報が１枚得られますので，これらをシステムに入力します．システムでは，これらの情報をそれぞれニューラルネットワークで処理し，低次元の空間にマッピングします．その情報を基に，ニューラルネットワークを用いて画像をつくり出します．この画像には，室内のどの場所でどのような属性の人物が発話しているかが表現されていますので，室内の大まかな様子を把握することができる，というわけです（**図２**）．このように，いったん入力を低次元空間にマッピング（エンコード）して，そこから高次元の情報を生成（デコード）する処理

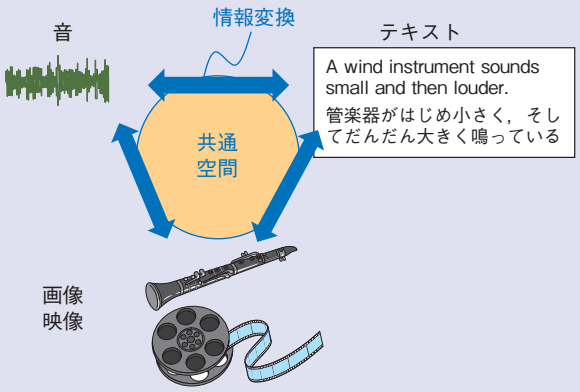


図１　クロスモーダル情報処理による情報変換の概念図

は，一般に「エンコーダ・デコーダモデル」と呼ばれ，入出力のペアを学習用データとして与えることで，深層学習によって構成することが可能です．

NTTコミュニケーション科学基礎研究所では，現在までにシミュレーション実験や実際の音を発する物体を使った実験を行って，一定の条件下で，どこに何があるかを画像として示すことが実際に可能であることを確認しています[(1)]．このような音から画像への変換は，これまで試みられたことがない新しい情報処理を提案したものとなりました．この技術が発展すると，カメラを置くことが望まれない場所やカメラがとらえきれない状況（物陰や暗闇など）での安全確認などにも応用できると考えています．

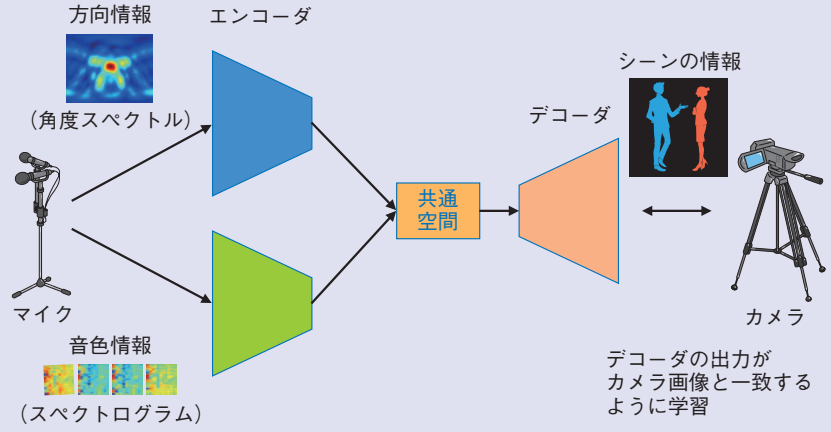


図２　音からの画像の生成

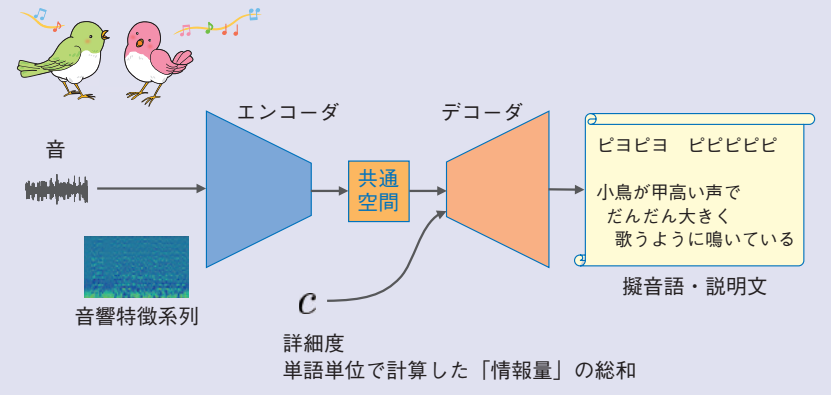


図３　音からの説明文の生成（条件付系列変換型説明文生成法）

⑵　物音を言葉で説明する

異種情報の変換のもう１つの例は，音からテキストへの変換です．音声認識システムを用いると話し言葉をテキストに変換できますが，これまでの音声認識システムでは，話し言葉以外の物音などを適切なテキストに変換することはできませんでした．これに対し私たちは，マイクで拾った音から，その音を表現する擬音語や，その音を記述する説明文を生成する技術を開発しました[(2)]．

条件付系列変換型説明文生成法（CSCG: Conditional Sequence-to-sequence Caption Generation）と呼ぶこの手法も，エンコーダ・デコーダモデルに基づいています（**図３**）．今度は系列から系列への変換（系列変換）を行います．まず，入力音響信号から抽出した特徴を時系列としてニューラルネットワークでエンコードし，低次元空間にマッピングします．次に，その情報からニューラルネットワークで音素系列（擬音語）または単語系列（説明文）をデコードします．

説明文の生成においては，どのような説明文を生成するのが適切かは場合によって異なり，唯一の正解を定めることはできません．例えば，「車が近づいている，危ない」といったように端的に短文で表現すべき場面もあれば，車種や車速などによるエンジン音の微妙なニュアンスの違いを詳細に表現したい，といった場面も考えられます．このような要請にこたえるため，

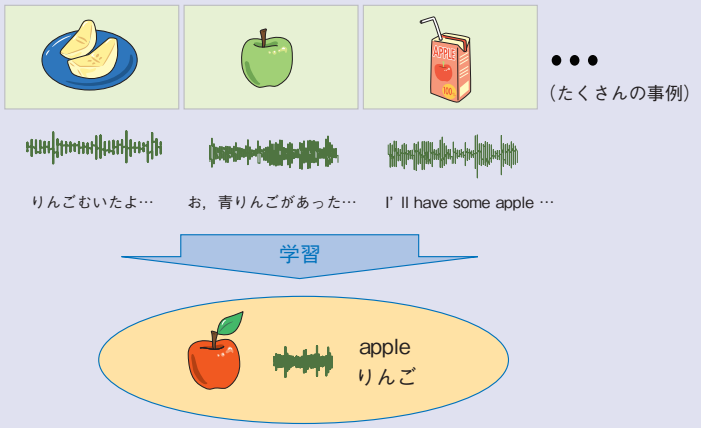


図 4　クロスモーダル情報からの概念獲得

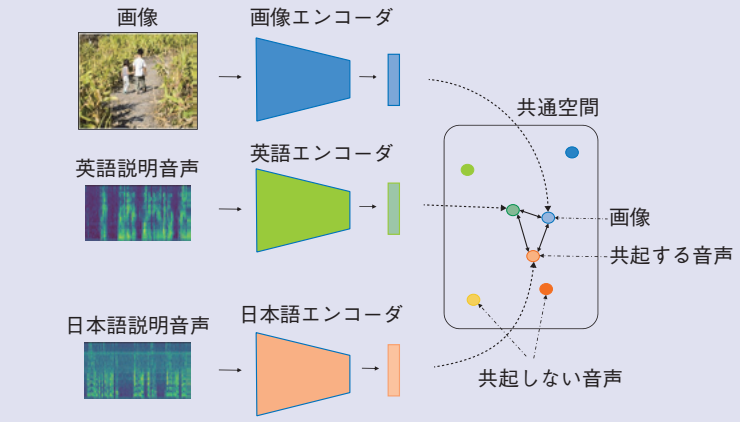
デコーダの働きを「詳細度」と呼ぶ補助入力で制御し，表現の詳しさ（説明文に含まれる単語の持つ情報量の和）を調節できるようにしました．小さな値の詳細度を指定すると端的な説明文を生成し，大きな値の詳細度を指定するほど，より具体的で，より長い説明文を生成するようになります．所定の条件における実験において，擬音語生成では人手による擬音語よりもむしろ受容度（あてはまっていると判断される割合）が高い擬音語の生成が可能であること，説明文生成や詳細度の制御も有効に機能すること，などを示しています．

本技術は，動画や実環境に対する字幕生成や，メディアの検索などに有効であると考えています．従来，音に対して「発砲音」「叫び声」「ピアノの音」などといったように既知のクラスラベルを与えることは試みられていました．しかし，画像の場合に比べても，音の場合には，音の信号と「何の音か」との対応が明らかであるとは限らず，「何かは分からない初めて聞くような音」に遭遇することは日常少なくありません．このような場合にはクラス分類だけでは有効性に限界があります．また，本技術では，音と説明文とが紐付くことにより，説明文による音の検索が可能となります．実際，共通空間においては音と擬音語や説明文との間の距離を直接測定することができ，擬音語や説明文を用いて音を検索することが可能です．このような場合，目的とする音のニュアンスを説明文で詳しく指定したい場合もあるでしょう．本技術を用いると，「車」「風」などといったクラスラベルだけではなく，音の高さや大きさ，変化の様子なども含めて，文字によって目的の音を指定することが可能になります．このような，音に対する説明文の生成も，私たちが世界で初めて提案した情報処理です．

概念獲得－未知の概念を自ら学習する

クロスモーダル情報処理のもう 1 つのメリットは，「共通空間」において異種情報の対応を見出すことで概念獲得が可能になることです．深層学習に必要とされる大量のデータの準備には，手間がかかったり，データの入手自体が難しかったり，クラスラベルの付け方を事前に設計することが難しかったりといった困難さを伴うことが少なくありません．そこで私たちは，メディア情報の中に含まれるひとまとまりのもの，つまり概念を自動的に獲得し，認識や検索に活用することをめざした研究に取り組んでいます．

異種のメディア情報の「共起」，つまり現実世界の中で，同じものに端を発する異種のメディア情報がランダムにではなく特定の関係性を持って現れることなどをうまく利用すると，人手でメディアデータどうしをペアリングすることなしに，共通空間を介したメディアデータのペアリングが可能になります．これを用いると，事前に「りんご」の画像とクラスラベルのペアを与えなくても，「皆がこの物体を指して“りんご”と言っているようだ．これは“りんご”というものなのだな」といった方式での学習が可能になるのです（**図 4**）．画像や音を見聞きするだけで賢くなる，というわけです．し



共起する情報を共通空間で近傍に配置するようにエンコーダを学習することで，音の各部分と画像の各部分との関連性を抽出できる

図5　画像および複数の言語情報の共通空間へのマッピング

かも周囲が“りんご”と言えばりんご，“apple”と言えば apple であると学習するといったように，周囲の人間の感じ方や振る舞い方を習得していきます．これは，私たち人間が，生まれてから成長するにつれて日常生活の中でさまざまなことを学んでいく過程に例えることができるでしょう．

私たちは，実際に，多数の写真に対して英語と日本語で何が写っているかを説明したもの（上記の，画像と音の共起を人工的に発生させたもの）を用いて，各言語における単語と，写真の中に写っている物体との対応付け（セグメンテーション）が可能であることや，画像を介した言語間の翻訳知識の自動獲得が可能であることを確認しています[3]（**図5**）．

今後の展開

本稿では，「クロスモーダル情報処理」が切り拓く新しい情報処理の最先端について，その一部を紹介しました[4]．これらの一連の研究に共通する考え方は，音，画像，テキストといったさまざまなメディア情報に対して，私たちの目や耳に触れる表層の表現形式と，その背後にある共通空間，つまり特定の表現形式には依存しない本来的な情報とを，分離して取り出してそれぞれを活用しようということです．これは多様な可能性を秘めた新しい情報処理の試みといえます．このような研究が発展すれば，私たち人間とともに暮らし，感じ方や振る舞い方を共有しながら，自ら学習していくAIも実現できそうに思われます．そのようなAIは，今よりもっと親しみを感じられるパートナーになり得るのではないでしょうか．

■参考文献

(1) G. Irie, M. Ostrek, H. Wang, H. Kameoka, A. Kimura, T. Kawanishi, and K. Kashino: “Seeing through sounds: Predicting visual semantic segmentation results from multichannel audio signals,” in Proc. ICASSP 2019, Brighton, U.K., May 2019.
(2) S. Ikawa and K. Kashino: “Generating sound words from audio signals of acoustic events with sequence-to-sequence model,” in Proc. ICASSP 2018, Calgary, Canada, April 2018.
(3) 大石・木村・川西・柏野・Harwath・Glass：“画像を説明する多言語音声データを利用したクロスモーダル探索，”信学技報，Vol.119, No.64, PRMU 2019-11, pp.283-288, 2019.
(4) Hot News: “NTTの「クロスモーダル」幼児のように世界を理解―生成AIの急激な発展で実現可能に，”日経エレクトロニクス，pp.14-15, 2019. 7.

柏野　邦夫

ロボットがたくさんの動画を見たり，周囲を見回しながら人の会話や環境中のさまざまな音を聞いたりするだけで，言語を覚え，世界を理解していくようになるのは，そう遠くない将来かもしれません．

◆問い合わせ先
NTTコミュニケーション科学基礎研究所
企画担当
TEL　0774-93-5020
E-mail　cs-liaison-ml@hco.ntt.co.jp

あなたの目の機能を気軽に楽しく測ります

気軽に楽しく自分の目の機能をセルフチェックできる仕組みの実現に向けた，２つの視覚測定テストについて紹介します．いずれも汎用的なタブレットデバイスで動作し，短い時間で目の機能を測ることができます．これらのツールを使ったセルフチェックが実現し，そのデータが多くの人から集められるようになれば，眼病における未病状態や，人の視覚機能の個人差の実態の解明につながります．

丸谷 和史（まるや かずし） /細川 研知（ほそかわ けんち）
西田 眞也（にしだ しんや）
NTTコミュニケーション科学基礎研究所

視覚の機能測定

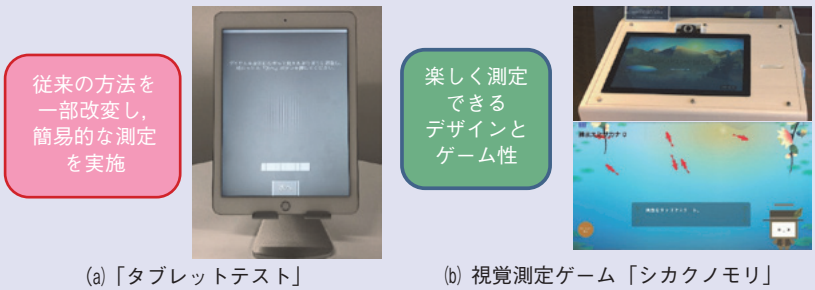
近年の社会の変化に伴って，視覚の機能測定は，より重大な課題になりつつあります．例えば，眼病は高齢になるにつれて発症率が高くなっていきます．高齢社会が訪れ，そして高齢化が進むと予測されている状況では，眼病を患う方の数は増加していくと考えられます．高齢化以外でも，視覚の機能測定は重要です．私たちはさまざまな表示装置に囲まれて暮らしています．それらの装置が自分自身の目に及ぼす影響については，未知の点が多くあります．さらに，近年では装置の種類も増え，選択の幅が広がっています．装置を選ぶときに，自身の視覚の特徴を知っておくことは重要です．また，装置設計の側でも，視覚機能の個人間でのばらつき，視覚多様性について，ある程度の知識が必要になると考えられます．

視覚機能測定は，普通病院などで行いますが，忙しい日常の中で，特に異常を感じていないときに病院で検査を受けることは実際には難しいでしょう．目の機能をより気軽に自分でチェックできる仕組みがあれば，より簡単に自分の特徴を知ることができます．しかし，これまで病院での検査や視覚科学の実験で使われてきた方法を，セルフチェックにそのまま適用しようとすると，さまざまな不都合が生じます．例えば，視覚検査では測定の精密さが要求されるので，それなりの時間がかかります．また，もともと検査員が操作する機器や測定キットを使うので，１人ではできないことがほとんどです．さらに，検査のための課題自体もあまり楽しいものではなく，目の機能に不安を持っていない人が自発的に利用する場面には向いていません．

NTTコミュニケーション科学基礎研究所では，視覚科学のためのさまざまな実験を長年にわたり実施し，データ取得のノウハウを得てきました．私たちは，これまで蓄積されたノウハウをセルフチェックに活用することを考え，視覚の特徴を気軽にチェックできるテストを作成しました．

視覚能力のセルフチェックのための２つのテストセット

私たちは，２つの新しい視覚機能の測定テストを提案しています（**図１**）．１つは従来の測定法を踏襲しつつ，タブレットデバイス上で実施可能とした簡易視覚テスト「タブレットテスト」です．もう１つは，デザイン，演出，タスク設定を工夫することで，ユーザのモチベーションを向上させ，日常的な反復利用を誘導する，ゲーム形式の視覚テストセット「シカクノモリ」です[(1)]．これらの２つは，いずれも，汎用的なタブレットデバイスで，専門家



(a)「タブレットテスト」　(b) 視覚測定ゲーム「シカクノモリ」

図１　２種の視覚測定テスト

が付き添っていない状況でも気軽に実施できるものをめざしています．

■「タブレットテスト」

「タブレットテスト」を作成するにあたり，私たちは，白内障，緑内障，加齢黄斑変性などの眼病の早期発見に将来利用していただく可能性も視野に入れて，共同研究者の神戸アイセンター病院・仲泊聡先生のご協力をいただきながら，複数のテスト項目を考えました．その基本となるのは，視力測定とコントラストに対する感度の測定です（**図2**）．簡易的な視力測定では，ランドルト環と呼ばれる「C」の形をした検査図形がどちらを向いているかをタッチで答えることで視力を簡易的に測定します（図2 (a)）．コントラストの感度測定では，さまざまな幅の白黒の縞模様がどのくらい薄くても見ることができるか（これを「コントラスト感度」と呼びます）を測定します（**図3** (a)）．さまざまな眼病で，コントラスト感度が低下することが知られており[(2)]，コントラスト感度を測定することで，自分の目の健康についてユーザ自身で知ることができると考えられます．また，さまざまな幅の縞に対するコントラスト感度のデータは，その人の基本的な視覚能力を知るうえでの重要な情報となります．「タブレットテスト」では，周辺部をぼかしたさまざまな縞模様で，ユーザ自身がその縞の濃さをぎりぎり見える値に調整することで，コントラスト感度を測定します（図3 (b)）．

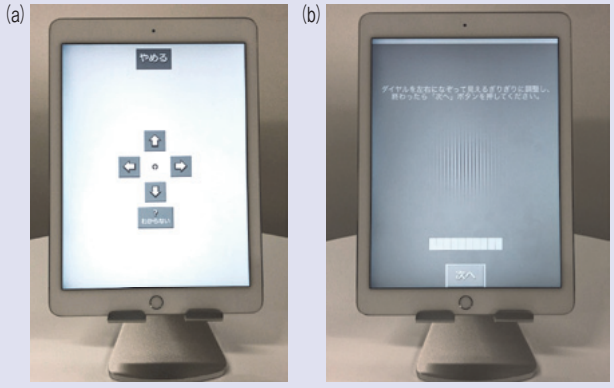
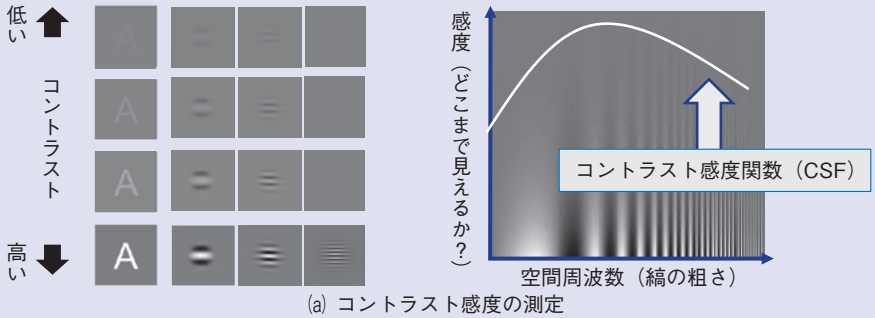


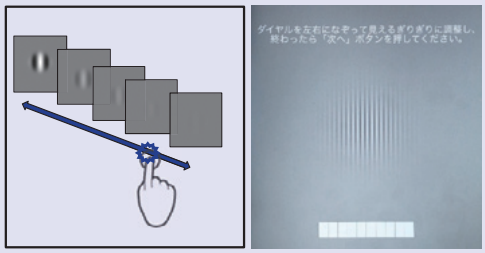
図2　「タブレットテスト」

■視覚測定ゲーム

もう1つのテストセットでは，ゲーム形式を導入し，これまでの視覚測定のイメージを大きく変えるようなグラフィックや測定方法自体のデザインを行いました（**図4**）．このテストセットでは，自分の視覚機能に特に異常を感じていない人でも，日常の生活場面

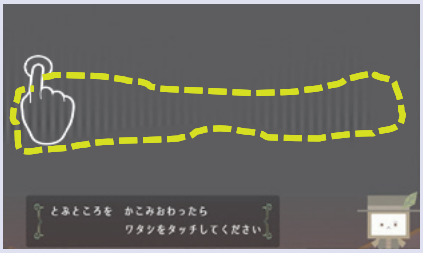


(a) コントラスト感度の測定



ユーザが縞のコントラストをギリギリ見える条件に調整

(b) タブレットテストでの測定法



ギリギリ見える境界をなぞって，1枚の画像で複数回判断

(c) 視覚測定ゲームでの測定法

図3　コントラストに対する感度の測定

図 4　視覚測定ゲーム

の中で少し空いた時間を使って，自分の視覚機能を確かめるといった用途を想定しています．本来はエンタテインメント性が必要ない課題に，ゲームの要素や形式を取り入れることによって，課題の理解が容易になる，課題へ取り組むモチベーションが向上する，課題により集中しやすくなる，などといったさまざまな良い効果が現れるとされています．実際に，視覚の分野でも，海外では主に子どもを対象とした弱視患者の視機能改善を目的としたゲームなどが開発されています(3)．日常的な視覚機能のチェックにもゲーム要素を取り入れることが有効であると考えられます．

このテストセットは4つのミニゲームから構成されています．それぞれ，コントラストの感度，視野の位置による感度のばらつき，周辺視野での文字認識の能力，運動する複数の物体の追跡能力のレベルについて，簡易的に測定できます．前者の2項目は，「タブレットテスト」にも含まれており，視覚の基本的な機能についてチェックをすることができます．後者の2項目は，視覚系の中でより高次の段階に位置する視覚認知処理も含まれる機能をチェックすることができます．コントラストの感度測定以外の3つの種目については，ビデオゲームと視覚機能の関係についてのこれまでの研究の中で，すでに重要であると考えられている機能を測定します．これらと，視覚科学の中でも重要と考えられているコントラストの感度測定（図3(c)）を組み合わせることで，広い範囲の視覚処理に対する機能のチェックを行うことができます．測定結果は，プレイ後の結果表示画面に現れるプレイスコア・グラフや，詳細なデータが記録されたQRコードなどで，知ることができます．QRコードは暗号化されており，専用のデータ解読ソフトウェアで，データ読み取り，データのグラフ表示ができます．

■テストの性能

私たちが作成したテストセットは，汎用的な機器での利用を前提として，Webブラウザ上で動作するアプリケーションとしました．測定は簡易的ですが，できる限り精確な測定を行うために，技術的な工夫を行っています．例えば，基本的な図形描画や時間の制御にJavaScript*1とWebGL*2を利用することで，精確な描画を実現しています(4)．さらに，一般的な機器における色階調は8 bit，256段階ですが，提案テストでは，コントラストの感度測定などの精度を上げるために，時空間ディザリング*3と呼ばれる手法によって，12 bit，4096段階までの色階調の表現を実現しています(5)．また，従来方法による測定のエッセンスを抽出し，日常的なセルフチェックにおいて重要度が低い手続きや測定条件を省略，逆にタブレットPCの特徴を利用した比較的新しい測定方法を導入することで，短い測定時間でも，できる限りテストの性能を向上させるように工夫しています．

私たちは，これらの工夫がテストの基本的な性能に反映されていることを確かめるために，実験を行いました．その結果，ゲーム化によってテスト利用時の楽しさが従来の実験法と比較して向上していること，テスト時間がおおむね3分程度に収まること，その一方で実験室での実験で数時間をかけて取得したデータと比較可能な精度でコントラストの感度を測定できることが分かりました(1)．また，ほかの種目でも，おおむね期待された精度での測定が行えることを確認しています．

簡易的な視覚測定セットが拓く可能性と今後の課題

私たちの作成した2種類の測定セットは異なる視点で作成されており，想定される利用状況も，それぞれ異なっ

＊1　JavaScript：Webページ作成向けのプログラミング言語．動的なWebページを実現するためなどに広く用いられています．
＊2　WebGL：2次元・3次元グラフィックス表示のためのWebブラウザ向けライブラリ．
＊3　時空間ディザリング：ある点の輝度，色などを，その点の時間・空間での近傍を含めた平均値で表現する技法．

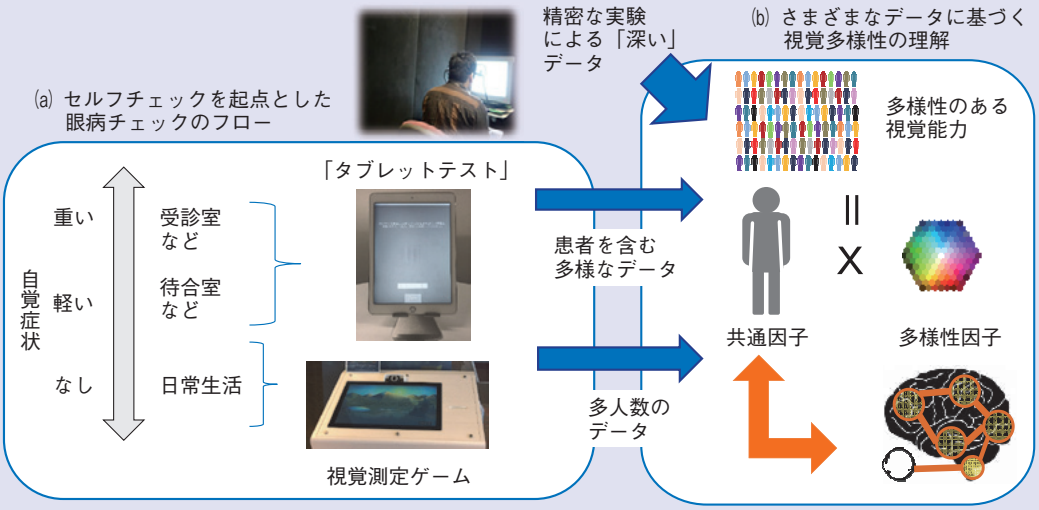


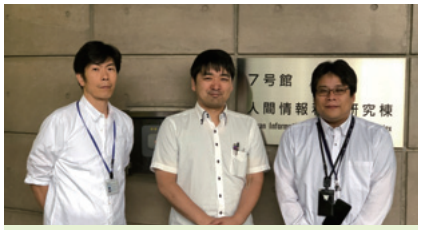
図5　複数のテストセットが拓く可能性

ています（**図5**(a)）．例えば，視覚測定ゲームは自宅を含む日常的な状況で短時間，繰り返し利用されることを想定しています．繰り返しの中で，ユーザは自分の視覚能力に基づくゲームスコアの範囲を知っています．そのスコアが継続的に低下したときには，自分の目に何らかの異常が起こっている可能性があります．その際には，より従来の検査に近い「タブレットテスト」を利用することで，測定を行うことが望ましいでしょう．そこでも異常が感じられる場合に，眼科を受診することになります．このような流れに沿って，提案したテストは，眼病の早期発見や病院外でのリハビリテーションに役立てていただけると考えています．

さらに，ここで提案しているテストセットは，視覚能力の多様性を研究するための方法として，視覚科学の研究に活用できる可能性があります（図5(b)）．汎用機器で気軽に視覚能力をチェックできる視覚測定ゲームからは，多くの健常者，あるいは軽度の異常を持った方々のデータが，病院の待ち受けなどに設置された「タブレットテスト」からは，眼病患者を含む多様な群からのデータが得られるでしょう．これらの多量・多様なデータをこれまでの視覚科学で蓄積してきた精密な実験による深いデータを合わせることで，視覚能力の多様性の解明と，多様性を生み出す因子の研究を進められます．この可能性を実現するためには，作成中のテストセットを実際に多くの人が触れられる状態にすることが必要です．これまでにも，共同でプロジェクトを進めている神戸アイセンター病院や，イベントなどでの試験を続けていますが，インターネットでの公開などを通じて，より多くの人に試していただける準備を進めています．

■参考文献

(1) K. Hosokawa, K. Maruya, S. Nishida, M. Takahashi, and S. Nakadomari: "Gamified vision test system for daily self-check," Proc. of IEEE GEM, 2019.
(2) A. Atkin, I. Bodis-Wollner, M. Wolkstein, A. Moss, and S. Podos: "Abnormalities of central contrast sensitivity in glaucoma," Am. J. Ophthalmol., Vol.88, No.2, pp.205-211, 1979.
(3) C. Gambacorta, M. Nahum, I. Vedamurthy, J. Bayliss, J. Jordan, D. Bavelier, and D. Levi: "An action video game for the treatment of amblyopia in children: A feasibility study," Vision research, Vol.148, pp.1-14, 2018.
(4) K. Hosokawa, K. Maruya, and S. Nishida: "Testing a novel tool for vision experiments over the internet," Journal of Vision, Vol.16, No.12, p.967, 2016.
(5) R. Allard and J. Faubert: "The noisy-bit method for digital displays: Converting a 256 luminance resolution into a continuous resolution," Behavior Research Methods, Vol.40, No.3, pp.735-743, 2008.
(6) http://www.kecl.ntt.co.jp/people/maruya.kazushi/index_ja.html



（左から）西田　眞也/ 細川　研知/ 丸谷　和史

自分自身の視覚の特徴を短い時間で，気軽に測定できるテストセットを作成しています．テストがどのようなものかを体験できるコンテンツを以下のHP(6)で公開していますので，そちらもぜひご覧ください．

◆問い合わせ先

NTTコミュニケーション科学基礎研究所
人間情報研究部
TEL　0774-93-5020
FAX　0774-93-5026
E-mail　cs-liaison-ml@hco.ntt.co.jp

錯覚　身体感覚　疑似歩行感

座っていても歩いているような疑似感覚の生成技術

NTTコミュニケーション科学基礎研究所では，座ったままの状態であたかも歩いたような感覚をつくり出す技術を開発し，この擬似的な歩行感によって私たちの脳内表現である自己の身体を取り囲む「身体近傍空間」を前方に拡張することを併せて明らかにしました．本稿では，椅子を上下に揺らし，足裏に振動刺激を与えることによって，実際に歩くことなく歩行したような感覚を生み出す手法を紹介します．

あめみや　ともひろ ※
雨宮　智浩

NTTコミュニケーション科学基礎研究所

VRにおける新たな歩行表現に向けて

近年，VR（Virtual Reality）ゴーグルと呼ばれる頭部装着型ディスプレイ（HMD）のような高性能でありながら低廉なデバイスの登場によりVR技術はゲームやエンタテインメントの分野で注目を集めてきました．さらに，ビジネス向け市場でも応用分野が広がりつつあり，医療分野での外科手術の訓練，生産現場での作業員教育，建設現場での安全教育など，さまざまな業界で注目されています．現在普及しつつあるVR体験ではHMDによる視覚への情報提示が中心ですが，日常生活の実体験では私たちは五感のあらゆる情報を身体を通じて接しているため，視覚だけでなく，複数の感覚への情報提示や，自己の運動感覚の生起が質の高いリアリティを生み出すためには不可欠と考えられています．

特にVR空間では利用者が歩いたり走ったりするような歩行・移動感覚の生成は大きな課題となっていました．実空間には空間の広さに制限があるため，広さの制限のないVR空間を歩き回るには工夫が必要で，例えば，トレッドミルのように歩いた分の移動量を相殺するような手法や，歩いている曲率や経路を気付かれないように調整する手法が数多く提案されてきました．こうした手法はVR空間を歩き回ることに有効ではあるものの，利用者が実際に歩行することを前提としたもので，空間的あるいは身体的な制約により歩行が困難な利用者に対して適用することができませんでした．そこで，NTTコミュニケーション科学基礎研究所では，利用者を実際に歩行させるという前提を見直し，座ったままの状態であたかも歩いたような感覚をつくり出す技術の開発に取り組んできました．本技術によって，例えば自宅のリビングに座ったまま，歩行したような感覚を移動範囲の制約を受けずに体験することができます．本稿では，身体的な多感覚刺激を用いた擬似的な歩行感覚の生成技術と，それを評価するための取り組みを紹介します（**図1**）．

※　現，東京大学大学院

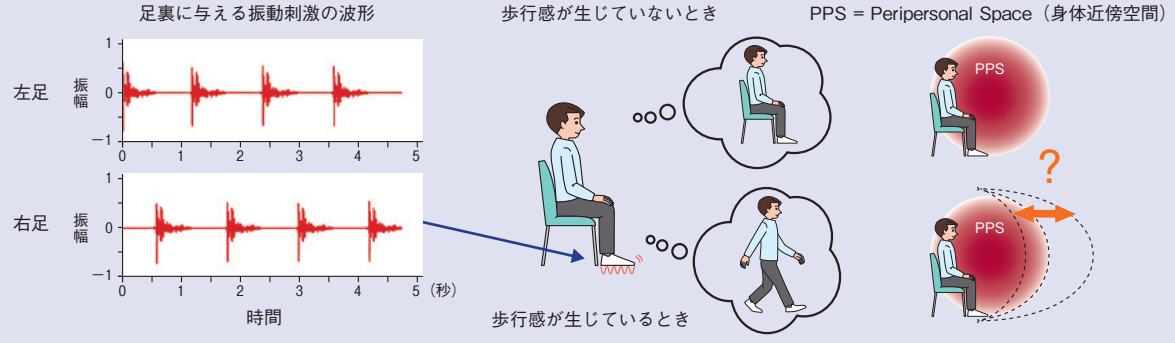


図1　足底への刺激による疑似歩行感と身体近傍空間の関係

足裏への振動刺激

座っている体験者に対して，視聴覚情報に加えて足裏に触覚刺激を受容したとき，歩行時に生じるような振動波形や歩行周期・タイミングといった特徴が一致する刺激を足底に与えることで歩行感覚に近い感覚が生じることを実験から明らかにしました．

足底は歩行時に地面との相互作用を行うインタフェースであり，足底からの情報から地面の状態や材質を知覚できることが知られています．また，新生児の自立歩行反射*1が知られているように足裏の触覚から歩行動作という一連の流れは生まれつき備わっているシステムでつくられるものと考えられます．このような点から本技術では足裏に着目し，適切な振動刺激を与えることで歩行運動を想起させ，歩行状態を擬似的に再現することを試みています．

具体的には，歩行時に実際に生じた振動を歩行音として記録し，カットオフ周波数120 Hzローパスフィルタ処理や増幅処理を経てボイスコイルモータ（振動子）で振動刺激として足裏に提示します（**図2**）．実験ではスリッパの踵部分に取り付けたボイスコイルモータによって提示しました．この歩行音の波形を正弦波に変えたり，歩行音の時間間隔を無作為化したりすると歩行感覚が生じないことも確認しました．

身体近傍空間の拡張

こうした刺激によって生成される歩行感は，体験者の主観評定に加え，私たちの脳内表現である自己の身体を取り囲む「身体近傍空間」*2の定義が変わることからも実際の歩行時に似た特性を持つことが確認されました．

身体に接近してくるような音を聞いているとき，その音源が身体近傍にあるときには聴触覚間の感覚相互作用によって身体上の触覚検出課題に対する反応時間が短くなることが知られています[(1)]．さらに，歩行中にはこの反応時間がさらに減少することが報告されています．本技術では，実際に歩行することなく，足裏への振動刺激のみによって反応時間がさらに減少することを世界で初めて明らかにしました[(2)]．

本技術の効果を検証するため，次のような実験を実施しました．実験参加者の課題はペンダント型装置内のボイスコイルモータから胸部に提示された触覚振動刺激（閾上の強度で150 Hzの正弦波）を検出し，できるだけ早くボタンを押すことを求めました．その際，装着したヘッドフォンからホワイトノイズの音源が直線上に等速運動で

*1　自立歩行反射：新生児にみられる反射の1つで，抱えられた状態で足の裏が平坦な地面に触れると，歩行するかのように両足を交互に踏み出す反射のこと．足裏への触覚刺激と歩行運動動作の間に生得的な関係があることを示唆しています．

*2　身体近傍空間：私たちの身体を取り囲む空間では他者との物理的あるいは社会的な相互作用が直接行われます．この空間は身体近傍空間と呼ばれ，この空間内では身体から離れた空間と異なる神経生理機構や知覚的機能が存在することが知られています．

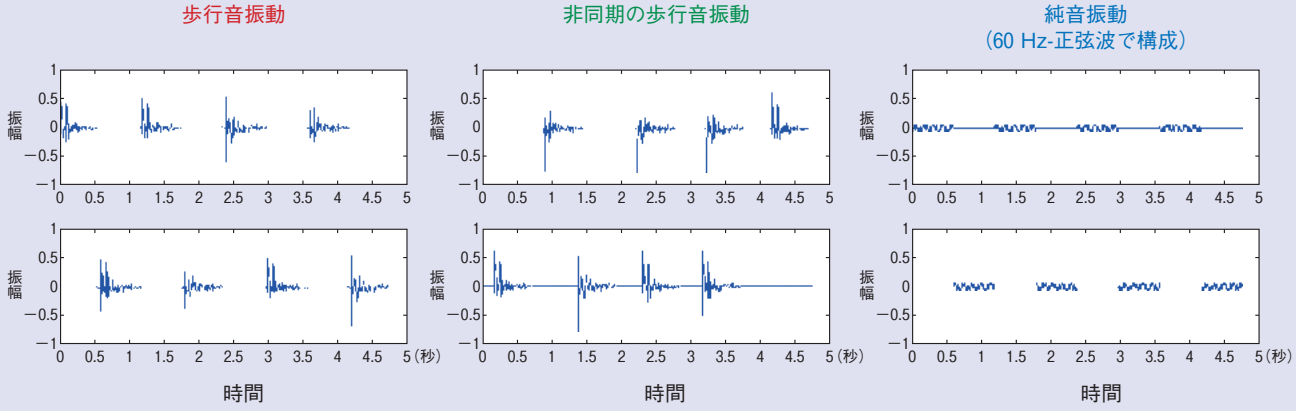


図2　実験で用いた足底への振動刺激パターン

接近することを模擬した音圧レベルが変化した音が提示されました．その接近音が身体近傍あるいは遠方にあるとき，胸部に振動刺激が提示されました．接近音が身体に近いときほど反応時間が小さくなると予想されます．スリッパの中敷きの踵の部分に組み込んだボイスコイルモータから，歩行音振動，歩行音振動を用いて左右の足へ非同期に歩行音振動を提示したり，歩行音振動を振幅一定の正弦波の振動に置き換えた音（純音）のいずれかの振動波形を提示して実験を行いました．

その結果，身体へ接近する音が身体に近いときほど胸部に提示された反応時間が小さくなることが確認されました．さらに，足への振動刺激によって受動的な状態で歩行感覚が生じたときに，歩行音振動条件でもっとも接近音が身体から遠いときでも反応時間が減少することを確認しました（**図3**）．このことは，胸部付近の身体近傍空間がより身体前方に拡張したことを示唆しています．さらに，歩行音振動条件では他の条件と比較して主観的に高い歩行感覚が得られていた（**図4**）ことから，歩行感覚の主観評定値の高さと身体近傍空間の前方への拡張との間に何らかの関係性があることが考えられます．

より高い臨場感をめざして

2019年5月に開催されたNTTコミュニケーション科学基礎研究所オープンハウス2019では，足底に着目した本技術に加え，モーションチェアと組み合わせて前庭感覚や固有感覚・触覚といった複数の感覚の同期提示によってさらに臨場感を高めた擬似的な歩行感を体験できる実演展示を行いました（**図5**）．私たちの先行研究では，歩行時に身体の揺動を与えたり，環境に応じて風や匂いを与えたりすることで，歩行体験や旅行体験を生み出せることを確認してきました[3]．今回の体験展示では，視覚情報に加えて，足裏への振動刺激と身体揺動という構成ながら，身体の揺動量やタイミングを工夫することによって，より高い歩行体験を生み出せることを確認しま

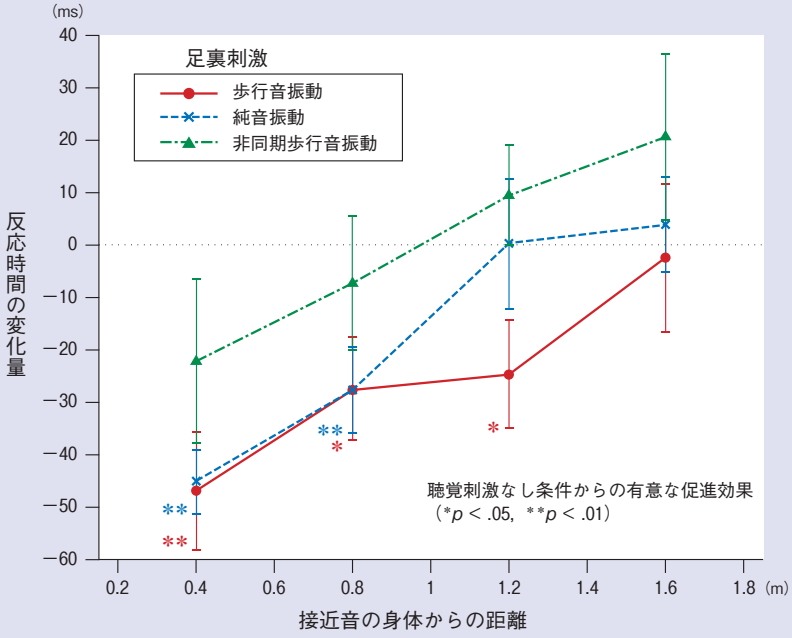


図3　足裏刺激ごとの触覚刺激に対する反応時間の変化

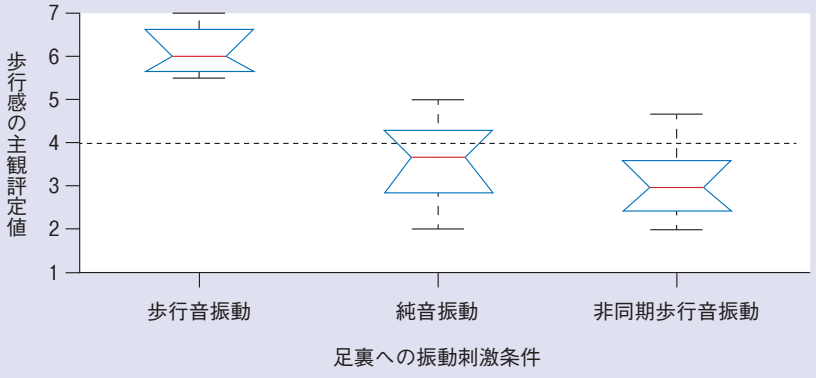


図4　足底刺激ごとの歩行感の主観評定値

図5　身体揺動と足底振動による擬似歩行感の生成

した．

今後の展開

本成果である視聴覚情報に加えて，前庭感覚および固有感覚・触覚に対する受動的な刺激の同期提示によって，座位のままで全身の歩行感覚を再現提示する研究は，調査した限りでは私たちの共同研究グループの研究を除いて過去に例がありません．これまでもNTTコミュニケーション科学基礎研究所では人間の知覚特性を活用した研究成果によりさまざまな課題を乗り越えてきました[(4)～(6)]．今回実現した技術について，今後，4D映画館やVRアミューズメント施設などでのVR空間内の歩行体験を高めるための要素技術として，本技術の応用の検討を進めます．さらに，歩く動作だけでなく，走る，スキップするなど多様な歩行感覚の表現をめざすとともに，体験者の足踏みなどの運動入力と連携して歩行感覚を生み出すための方法論の確立をめざします．

人間の視覚・聴覚特性が映像装置や音響装置の設計において考慮されているように，触覚や力感覚をはじめとする五感情報通信が将来実現された場合も，人間の知覚側から情報提示装置の設計指針を規定することが重要になると考えられます．そのため，人間の感覚知覚機序の解明を進めながら，その人間の知覚特性を利用してさらなる五感インタフェースの研究へと転換できるような礎としての基礎研究を進めていきたいと考えています．

本研究の一部は，首都大学東京，豊橋技術科学大学との共同研究によるもので，平成30-32年度 文部科学省 科学研究費補助金 基盤研究(B) 18H03283「擬似身体移動感の定量的評価法の開発とそれを用いた多様な移動感の生成手法の確立」(研究代表者：雨宮智浩)の助成を受けました．

■参考文献

(1) J.-P. Noel, P. Grivaz, P. Marmaroli, H. Lissek, O. Blanke, and A. Serino: "Full body action remapping of peripersonal space: the case of walking. ," Neuropsychologia, Vol.70, pp.375–384, 2015.
(2) 雨宮・池井・広田・北崎："歩行を模擬した足底振動刺激による身体近傍空間の拡張," 日本バーチャルリアリティ学会論文誌, Vol.21, No. 4, pp. 627-633, 2016.
(3) 池井・広田・阿部・雨宮・佐藤・北崎："身体的追体験の概念の提案と一部機能の試験実装―多感覚・運動情報提示による歩行・走行体験の共有," 日本バーチャルリアリティ学会論文誌, Vol.24, No. 2, pp.153-164, 2019.
(4) 雨宮・安藤・何："五感インタフェースによるノンバーバルコミュニケーション," NTT技術ジャーナル, Vol. 19, No. 6, pp. 35-37, 2007.
(5) 雨宮・高椋・伊藤・五味："指でつまむと引っ張られる感覚を生み出す装置「ぶるなび3」," NTT技術ジャーナル, Vol.26, No.9, pp.23-26, 2014.
(6) 雨宮："触覚・身体感覚の錯覚を活用した感覚運動情報の提示技術," 基礎心理学研究, Vol. 36, No.1, pp.135-141, 2017.

雨宮　智浩

五感を通じて体験する技術を先鋭化させるために，人間の特性を理解し，その知見を活かすことを心掛けて研究を続けています．今後も安心・安全・快適な社会に向け，知の体系化，そして革新的な技術の創出をめざします．

◆問い合わせ先
NTTコミュニケーション科学基礎研究所
企画担当
TEL　0774-93-5020
FAX　0774-93-5026
E-mail　cs-liaison-ml@hco.ntt.co.jp

文脈を理解して話す雑談対話システム

人にとって何気ない雑談であっても，同じような雑談を対話システムが行うにはたくさんの難しさがあります．近年の雑談対話の研究によって，雑談対話の最初の課題であった幅広い話題に対する応答ができるようになってきました．しかしながら，文脈をふまえた応答は難しく，「このシステム分かってない」とユーザに感じさせることが多々ありました．本稿では，文脈を理解して話す雑談対話システムの取り組みについて紹介します．

成松 宏美 /杉山 弘晃
水上 雅博 /有本 庸浩
宮崎 昇

NTTコミュニケーション科学基礎研究所

雑談対話システムの実現をめざして

マイデイズをはじめとするスマートフォン上のエージェントやAIスピーカなどの普及に伴い，人と機械の対話が増えてきています．現在商用として使われている対話システムの多くは，主に，「Aさんに電話して」「今日の天気を教えて」，などのタスク実行を目的としていますが，雑談ができる対話システムにも期待が高まっています．雑談をすることには多くの効果があるといわれており，記憶の整理や人のコミュニケーションスキルの向上にも役立つと期待されています．NTT コミュニケーション科学基礎研究所では，早くから「雑談」に着目をして研究を行ってきました．

雑談を行う対話システムでは，先述のタスクを行う対話システムと異なり，ユーザの発話の話題が幅広く，ユーザの発話を事前に想定して設計することができないという難しさがあります．例えば，レストラン予約などのタスクを行う場合には，予約日時や予約者の名前と電話番号など予約をするうえで必要な情報は決まっており，ユーザの発話に含まれ得る情報をあらかじめ想定することが可能です．一方で，雑談対話では，ユーザの発話に含まれる情報を想定しておくことは難しく，ユーザのあらゆる発話に対して適切に応答することが困難でした．

私たちの所属する研究グループでは，幅広い話題に応答できるようにするために，質問と応答，発話と応答，発話と質問などのさまざまな応答ペアを多量に用意し，機械学習の訓練データとして用いたり，文間類似度により発話選択をしたりする手法に取り組んできました．これまでの研究成果により，一問一答ベースでは，ユーザの発話に対してある程度近い応答はできるようになってきました．

しかしながら，人と同じように対話できる相手になるためには，相手の発話に合わせた適切な応答や，文脈に整合した適切な応答ができなければなりません．本稿では，これらの問題に対する，私たちの最新の研究成果を紹介します．

一問一答ベースの対話システムの問題点

従来の一問一答ベースの対話システムでは，直前のユーザ発話に対して，多量に用意した発話例から近い応答を返すという戦略でした[(1)]．そのため，違和感のある発話や，それ以前の対話を聞いていないような発話をすることがあり，数分の対話でも「分かってくれてないな」とユーザに感じさせてしまうという課題がありました．例えば，従来のシステムでは，ユーザとの対話がしばしば次のようになります（**図1**）．この対話では，初めにユーザが「夏休みにたこ焼き食べた」と言ったのにもかかわらず，4発話目にシステムが「いつ行ったんですか？」と聞いてしまい，ユーザに「さっき夏休みって言ったのに，分かってないな…」と思わせています．また，「夏休みは避暑地がいい」という発話も，これまでのたこ焼きの話題から急に離れてしまうため，システムがなぜその発話をしたのか分からず，ユーザを困惑させています．こうした，①直前までの対話の内容と整合しない発話や，②根拠のない発話によって，ユーザは，「システムは分かってないな」や「このシステムは何が言いたいか分からない」と感じてしまい，システムと対話するのを諦めてしまう原因となります．これでは，コミュニケーションする相手として良いといえないどころか，ユーザに「対話できる相手」として認めてもらえず，使ってもらえなくなります．

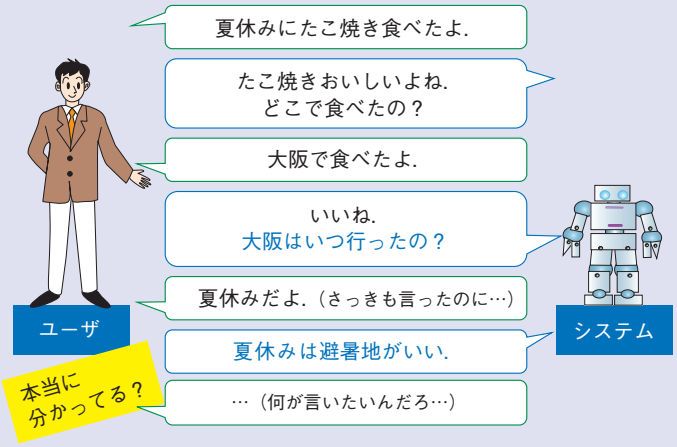


図１　一問一答ベースの対話システムとユーザの対話例

「対話できる相手」になるために

「対話できる相手」として認めてもらうためには，少なくとも前述した問題を解消する必要があります．これは，心理学者のグライスが提唱した対話の成立条件[(2)]でも言われています．関連のないことを言ってはいけない（関連性の公準），根拠のない適当なことを言ってはいけない（質の公準）が挙げられており，①文脈に整合しない発話や，②根拠のない発話は対話破綻を導くとされています．そこで，２つの問題に対して，文脈に整合した発話および根拠のある発話を行うために，私たちは，「文脈の理解」および２つの発話生成「文脈に整合した発話生成」「根拠に基づく発話生成」に取り組みました．以降では，それぞれに対する私たちのアプローチを紹介します．

文脈の理解

文脈となる現時点以前の対話をどのように理解し，文脈情報として保持したら良いでしょうか？

私たちは，ユーザの体験は5W1H+感想で表せることが多い点に着目し，5W1H+感想の情報を文脈として理解し，利用する方法を考えました．5W1Hのフレームは非常にシンプルですが，人とのコミュニケーションやカウンセリングの対話においても5W1Hの質問をしていく戦略がとられており，いろいろなシチュエーションで共通して使えるフレームです．例えば，旅行の話をするときにも，美味しいものを食べた話をするときにも，「どこに行ったの？」「いつ行ったの？」「どうだった？」などは自然な流れで出てくる質問だと思います．

では，5W1H+感想の情報はどのように理解したら良いでしょうか？

5W1Hのうち，時間や場所に関する情報は，従来から固有表現抽出技術で抽出の対象とされてきていました．例えば，「昨日，東京に行ったよ」という文が与えられた場合，「昨日」は時間，「東京」は場所の固有表現として抽出されます．主に，固有名詞や日付・時間に特有の表現などが抽出の対象となっています．でも，人の自由発話に含まれる時間や場所の情報は，これだけで足りるでしょうか．私たちが，実際の人どうしの雑談対話を収集し，人が「時間」や「場所」として理解するフレーズを調べた結果，固有表現だけではないフレーズが多くを占めていることが分かりました（場所フレーズの場合は約７割）．

そこで，ユーザの自由発話に含まれる5W1H+感想に該当するフレーズを抽出する，フレーズ抽出器を構築しました．フレーズ抽出器には，固有表現抽出でも有効とされている系列ラベリング手法を用います．代表的な手法は，CRF[(3)]ですが近年ではディープニューラルネットワークを用いた手法も提案されています．これらの手法に対して，人どうしの雑談対話中で5W1H+感想の各項目として人が理解するフレーズに対して人手でアノテーションを行い，学習させました[(4)]．

この結果，正式名称でなくとも「京都駅近くの公園」などが場所としてとれたり，「たこ焼きを食べた」をWhat項目としてとれたり，文脈に必要な情報をユーザの発話から新たに抽出できるようになりました．場所フレーズを例に，従来の固有表現抽出器とその抽出結果を比較すると，**表**のように，固有名詞を含むフレーズや，一般名詞などが抽出できるようになっていることが分かります．本技術を用いて，対話中に現れたユーザの5W1H+感想の情報を埋めていくことで，文脈を理解することができるようになりました．

表　ユーザ発話に含まれる場所フレーズの例と抽出結果の比較

ユーザ発話（赤字:場所フレーズ）	従来技術の抽出結果	本技術の抽出結果
夏休みにイタリアに行きました.	イタリア	イタリア
京都駅近くの公園でお花見したよ.	京都駅	京都駅近くの公園
暇なときは電気屋に行きます.	（なし）	電気屋

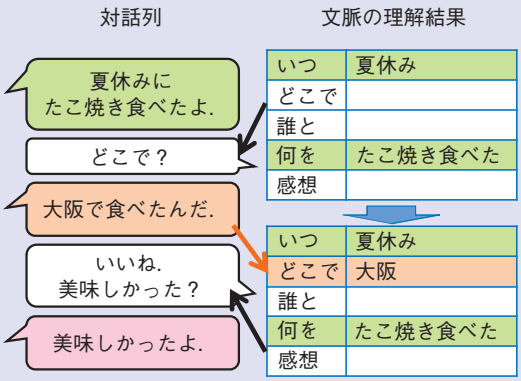


図 2　文脈の理解結果に基づく質問生成の例

文脈に整合した発話生成

前述の文脈理解の結果を用いることで，文脈に整合した質問や発話の生成が容易になります．例えば，5W1H+感想の情報を引き出す対話であれば，従来の技術で起きていた直前にユーザが話したことを質問してしまうことも抑制できるようになります（**図 2**）．また，対話中に「夏休みに旅行に行った」と「大阪観光した」という発話があれば，「時間：夏，場所：大阪」という情報から 2 つを関連付けて，「海遊館とか行った？」という関連する質問や，「夏の大阪は暑いですよね」という文脈に整合した発話を生成することができます．これは，従来手法で直前のユーザ発話（大阪観光した）のみから生成され得る「大阪には道頓堀がありますね」という発話よりも，より文脈に整合した発話となっていることが分かると思います．

根拠に基づく発話生成

文脈に整合するだけでは，根拠のある発話になることは保証されません．そこで，私たちは，システムが発話を行う際にその根拠となるような情報を付け加えて提示するアプローチを提案しました．ここでは質問と共感発話を例にアプローチを紹介します．

1 番目は，システムが質問をする際に，なぜ質問したかの理由を加える方法です．システムが「夏でも楽しめますか？」と質問をする際には，「私は夏休みに行けたらと思っているので，参考にさせてもらえたらなと思って」のように，質問した理由や根拠となる情報を付け加えます．これにより，システムがなぜその質問をしたのかをユーザに伝えることができます[5]．

2 番目は，システムが共感や感想を述べる際に，なぜそう思うのかの根拠情報を付け加える方法です．「たこ焼きは美味しいですね」と伝える際に，「私も大阪でたこ焼き食べましたよ．熱々だったので美味しかったです」というように，なぜその感想を抱いたかの理由や根拠となる情報を付け足します．例えば，システム自身の知識として，**図 3** のような構造を持っておき，それを発話のフォーマット「私も[場所]に行って[何をした]．[感想根拠]ので[感想]」に当てはめることで，根拠を含めた発話を生成することができます．これにより，単に「たこ焼きは美味しいね」というよりも，システムが自身の体験[6]や知識に基づき「本当にそう思って」共感している感じを与えられます．

先述の「文脈の理解」および「文脈に整合した発話生成」と本技術を組み合わせることで，文脈に整合した発話にさらに根拠を付け加えることが可能となりました．これにより，**図 4** のような「分かってくれる」対話が可能となります．

今後の展開

今回の取り組みにより，対話システムは，文脈を理解し，文脈に合わせた質問や根拠のある発話の生成ができるようになりました．これは，「このシステム分かってないな」とユーザに思われていた対話システムを，「理解して」対話できるシステムに変える大き

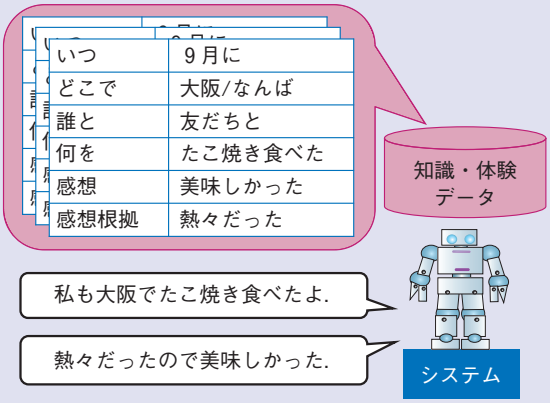


図3　システムの知識・体験データを根拠とした発話生成

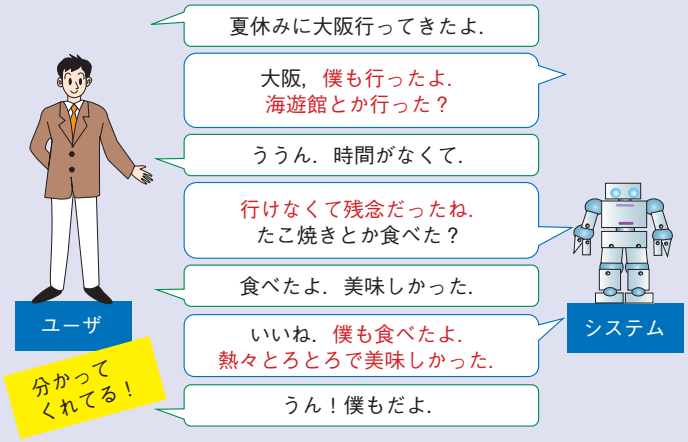


図4　本取り組みにより実現できるシステムとユーザの対話例

な一歩であると考えます．自分の話を理解して話してくれるシステムがいたら，いろいろな話をしたり相談したりするようになるでしょう．コミュニケーショントレーニングや相談などさまざまなシーンでの対話システムの活用促進にもつながると考えています．

しかしながら，これらの実現には，対話の流れをうまく設計し，システムの知識となるデータを人手で作成するなどの，手間とノウハウが必要であり，誰でも簡単に同様のシステムがつくれる状況には至っていません．

今後は，今回人手で作成したようなデータをWebや人との対話を通して自動的に作成する手法にも取り組んでいきます．

■参考文献

(1) 杉山・東中・目黒："気軽に雑談できるシステムの実現をめざして," NTT技術ジャーナル, Vol.28, No.9, pp.16-20, 2016.
(2) H.P. Grice: "Logic and conversation," Syntax and Semantics, Vol.3, Speech Acts, pp.41-58, 1975.
(3) J. Lafferty, A. McCallum, and F.C.N. Pereira: "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," Proc. of ICML 2001, pp.282-289, June 2001.
(4) H. Narimatsu, H. Sugiyama, and M. Mizukami: "Detecting Location-Indicating Phrases in User Utterances for Chat-Oriented Dialogue Systems," Proc. of LACATODA 2018, pp.8-13, July 2018.
(5) 杉山・成松・水上・有本："文脈に沿った発話理解・生成を行うドメイン特化型雑談対話システムの実験的検討," 人工知能学会 言語・音声理解と対話処理研究会（SLUD）第84回研究会（第9回対話システムシンポジウム）, 2018.
(6) M. Mizukami, H. Sugiyama, and H. Narimatsu: "Event Data Collection for Recent Personal Questions," Proc. of LACATODA 2018, July 2018.

（左から）杉山　弘晃/ 有本　庸浩/
成松　宏美/ 水上　雅博/
宮崎　昇

人は相手の理解状況に合わせて情報提示の方法を変え，円滑なコミュニケーションをしています．今後も，システム自身の理解力向上に取り組みつつ，対話相手の理解度に応じた柔軟な応答も検討していきます．

◆問い合わせ先

NTTコミュニケーション科学基礎研究所
協創情報研究部
インタラクション対話研究グループ
TEL　0774-93-5020
FAX　0774-93-5026
E-mail　cs-liaison-ml@hco.ntt.co.jp

限界まで効率良くメッセージを送れます ——シャノン限界を達成する通信路符号

本稿では，通信効率の限界（シャノン限界）を達成する実行可能な符号化技術CoCoNuTSを用いて構成した通信路符号（誤り訂正符号）を紹介します．本技術により既存の方法よりも効率の良い通信が実現できます．

村松（むらまつ） 純（じゅん）

NTT コミュニケーション科学基礎研究所

通信路符号

通信を行ううえでは，雑音のある環境下でも正しくメッセージ（情報）を伝える必要があります．これを実現する技術は「通信路符号」あるいは「誤り訂正符号」と呼ばれており，光通信や無線通信に限らず，計算機の内部やハードディスク・光ディスク等の記録装置，スマートフォン等で情報を読み取るための二次元コード等に応用されています．あらゆる通信機器の中に入っているといっても過言ではありません．

雑音のある環境（通信路）が与えられたとき，正しくメッセージを伝えることができる効率には限界があります．このような通信効率の限界は，1948年にこれを発表した計算機科学者シャノンにちなんで「シャノン限界*1」と呼ばれています．しかしながら，シャノンが提案した符号は膨大な計算量を必要としていたため，その実行は困難でした．実行可能なシャノン限界を達成する符号の構成は，シャノンが創始した情報理論の70年にわたる課題です．

その後，シャノン限界を達成する実用的な符号としてLDPC（Low Density Parity Check：低密度パリティ検査）符号*2などが開発され，近年の第5世代移動通信システム（5G）に実装されています．しかしながら，これらの符号がシャノン限界を達成するのはある特殊な通信路に限られており，一般の通信路では限界を達成できません．

研究の成果

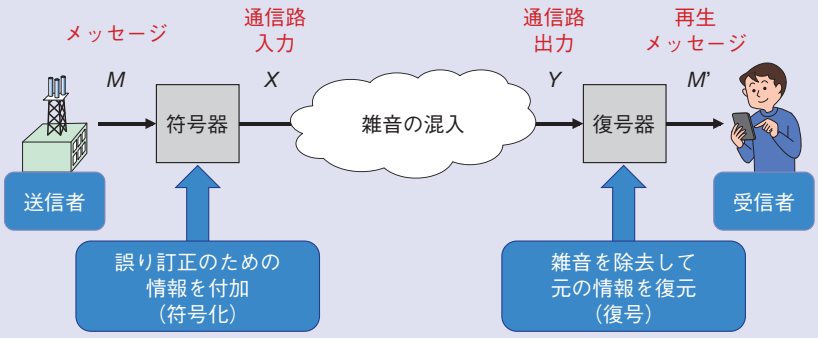
NTTコミュニケーション科学基礎研究所では，シャノン限界を達成する符号化技術CoCoNuTS（Code based on Constrained Numbers Theoretically-achieving the Shannon limit：拘束条件を満たす系列に基づくシャノン限界を達成する符号）*3を開発しました．本技術を用いることにより，通信路符号だけでなく，情報源符号や情報理論的安全性を持つ暗号などの通信のあらゆる問題に対して，限界を達成と実行可能性を両立させる符号を構築できます．今回は，本技術を通信路符号へ応用することにより，それがシャノン限界を達成できることを数学的に証明しました(1)～(3)．また，シミュレーション実験により，従来のLDPC符号ではシャノン限界を達成できなかった通信路に対して，提案法が従来法を超える性能を持つことを確認しました．

技術のポイント

■通信路符号（誤り訂正符号）が実現する通信システム

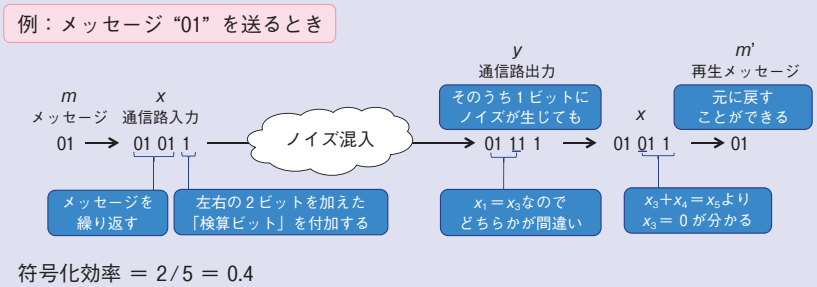
通信路符号が実現する通信システムを**図1**に示します．ここでは通信会社の基地局がメッセージを送信する送信者となり，スマートフォンを持ったユーザが受信者となっています．最初に符号器は送信したいメッセージMを符号化して通信路入力Xへ変換します．変換された信号は電波に変換（変調）されて送信されますが，電波を送受信して通信路に出力Yを得る際に雑音が混入することを想定します．復号器は通信路出力Yから元の再生メッセージM'を復元します．ここで，正しい通信とは，メッセージと再生メッセージが同一（$M=M'$）であることを意味しています．そこで，メッセージと再生メッセージが異なる（$M\neq M'$）事象の確率を「復号誤り確率」と定義します．この値が小さいほど性

*1 シャノン限界：通信路符号（誤り訂正符号）の文脈においては「通信路容量」という名称として知られているものです．
*2 LDPC符号：「パリティ検査行列」と呼ばれる低密度行列（成分のほとんどが0の行列）を用いて高速な復号を行います．
*3 CoCoNuTS：「拘束条件を満たす乱数生成器」を用いることにより限界を達成する通信を実現することから名付けました．



復号誤り確率＝（$M \neq M'$ になる確率）
通信効率＝（メッセージシンボル数）／（送信信号数）

図 1　通信路符号が実現する通信システム



符号化効率 ＝ 2/5 ＝ 0.4

図 2　符号化の具体例

能が良いことになります．一方で，符号化レートをメッセージシンボル数と送信信号数の比と定義します．この比が大きいほど通信効率は高くなり，高速な通信が可能になりますが，通信効率を大きくし過ぎると復号誤り確率を 0 に近づけることができなくなります．この通信システムでは，復号誤り確率が 0 に限りなく近いような符号化・復号化で，可能な限り大きな符号化レートを実現することをめざします．

■符号化の具体例

メッセージ “01” を送信することを例にして，符号化の具体的な方法を**図 2** に示します．符号器はメッセージ “01” を符号化して通信路入力 “01011” を求めています．この符号器では，メッセージを 2 回繰り返した後でメッセージの左右の 2 ビットの加えた「検算ビット」（排他的論理和）を付加するという操作を行っています．この例では，受信時にノイズが混入して “01111” という通信路出力が得られました．復号器は符号化のルールからノイズが混入した位置を特定して，再生メッセージ “01” を得ます．この手続きが復号と呼ばれるものです．今回は正しいメッセージと再生メッセージが一致して正しい通信が行われましたが，雑音によっては正しくメッセージが再生できない場合もあります．この例では，2 ビットのメッセージを 5 ビットの通信路入力へ変換したので，符号化レートは 2 / 5 ＝0.4になります．この例にあるメッセージの繰り返し回数や検算ビットを増やすことにより，より多くの雑音の位置を特定できるようになりますが，送信信号数が増えるため符号化レートは小さくなってしまいます．高速な通信のためには，雑音の位置を高い確率で特定できて，かつ符号化レートを可能な限り大きくできることが重要です．

■シャノン限界

図 3 はシャノン限界（通信路容量）を説明しています．図 1 では，符号が満たすべき条件として，復号誤り確率が 0 に限りなく近いことを要請しました．一方で，符号化レートが大きい符号ほど通信効率が良いことを説明しました．シャノン限界は復号誤りが 0 に限りなく近い符号の符号化レートの限界です．シャノンは符号化レートの限界が図 3 に示されている式に等しいことを示しました．この限界は送信信号数を十分に大きくとることによって達成できます．シャノン限界を超えた効率を持つ符号を設計することは理論上不可能であり，もしもこの限界を達成する方法が実現できれば，理論的

にはこれ以上の性能向上が見込めないことになります．なお，シャノン限界を達成するためには，通信路入力分布Pを最適化する必要があります．

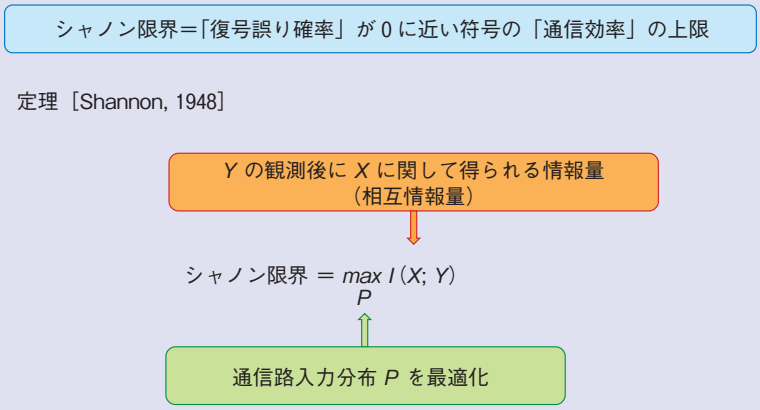


図 3　シャノン限界

■提案法CoCoNuTSの技術ポイント

CoCoNuTSを用いた通信路符号の構成を**図 4** に示します．提案法では，2 つの疎行列（成分のほとんどが 0 の行列）A，Bとベクトルcを用いて構成しています．さらに符号器と復号器に現れる写像 $f_{A,B}$，g_A（図 4 の赤枠の部分）として，後述の「拘束条件を満たす乱数生成器」を用いて最適な通信路入力分布Pを実現することにより，シャノン限界を達成することができるようになりました．これは図 4 に示されている定理を証明することによって理論的に保証されます[(2), (3)]．

従来法であるLDPC符号では，生成行列を用いて，メッセージと通信路入力を対応させます．このため，得られる通信路入力の分布は一様分布に近いものになります．したがって，図 3 で

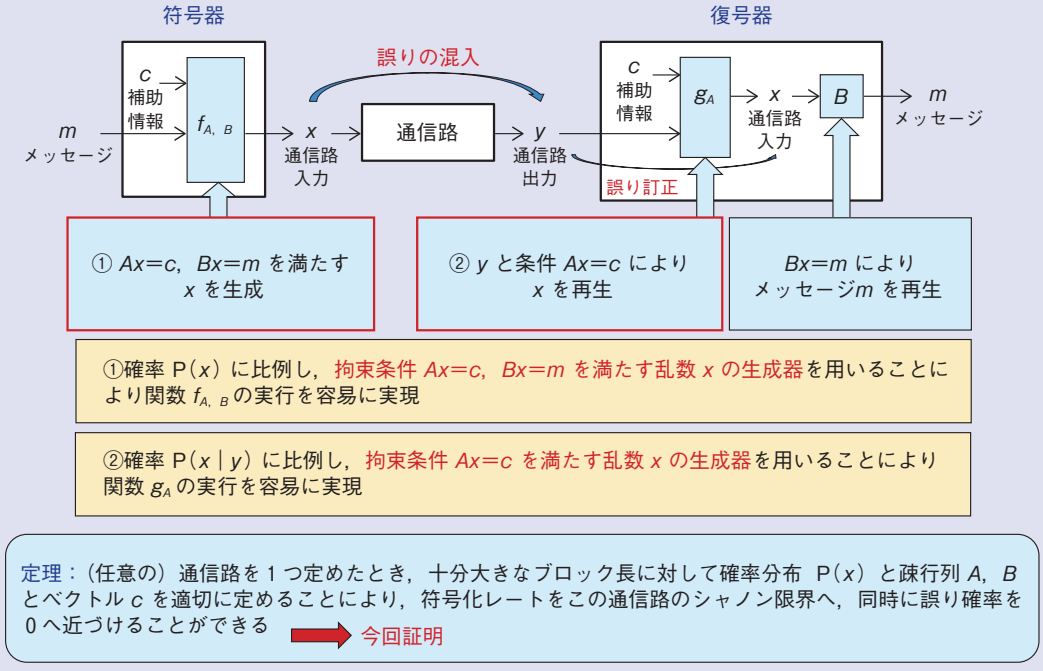


図 4　提案法CoCoNuTSの技術ポイント

示したシャノン限界の式で，通信路入力分布Pが一様分布のときに最大値（max）を達成していれば，LDPC符号はシャノン限界を達成しているといえます．逆に，最大値を達成する通信路入力分布Pが一様分布ではない場合はシャノン限界を達成できないことが分かります．

■拘束条件を満たす乱数生成器

図4の①，②では，「拘束条件を満たす乱数生成器」を用いています．これは，方程式で与えられた拘束条件を満たす系列xを与えられた確率に比例して生成させるものです．これによって，写像$f_{A,B}$とg_Aの実現が容易になります．ここで高次元の乱数xを直接生成する代わりに，系列の成分（一次元）の乱数の確率分布を求めて逐次的に発生させることにより，計算の実行が可能になりました[(2)]．

■シミュレーション実験

シミュレーション実験で提案法とLDPC 符号を比較した結果を**図5**に示します．グラフの横軸は符号化レートを示しており，右にあるほど性能が良いことになります．グラフの縦軸は復号誤り確率を示しており，下にあるほど性能が良いことになります．グラフは提案法がLDPCよりも良い性能を持つことを示しています．例えば，縦軸の誤り確率10^{-4}で両者を比較したとき，提案法は符号化レートで0.03，情報量に換算すると2000ビット当り60ビットの情報を多く送信できることが分かります．

符号化レートを固定して両者を比較したものを**図6**に示します．図5にある実験条件で符号化レート0.6の符号化を1200回繰り返したときに，復号に失敗した場合を黒点で示して頻度を可視化しました．LDPC符号では7回の復号誤りを観測したのに対して，提案法は一度も復号誤りを観測しませんでした．この結果から，同じ雑音環境で同じ量のメッセージを送信した場合は，提案法は従来法に比べてより信頼性が高いことが分かります．

提案法と従来法を実際の通信に近い状況で実行した場合の比較結果を**図7**に示します．実際の通信は復号誤りがないことを想定して画像を圧縮し

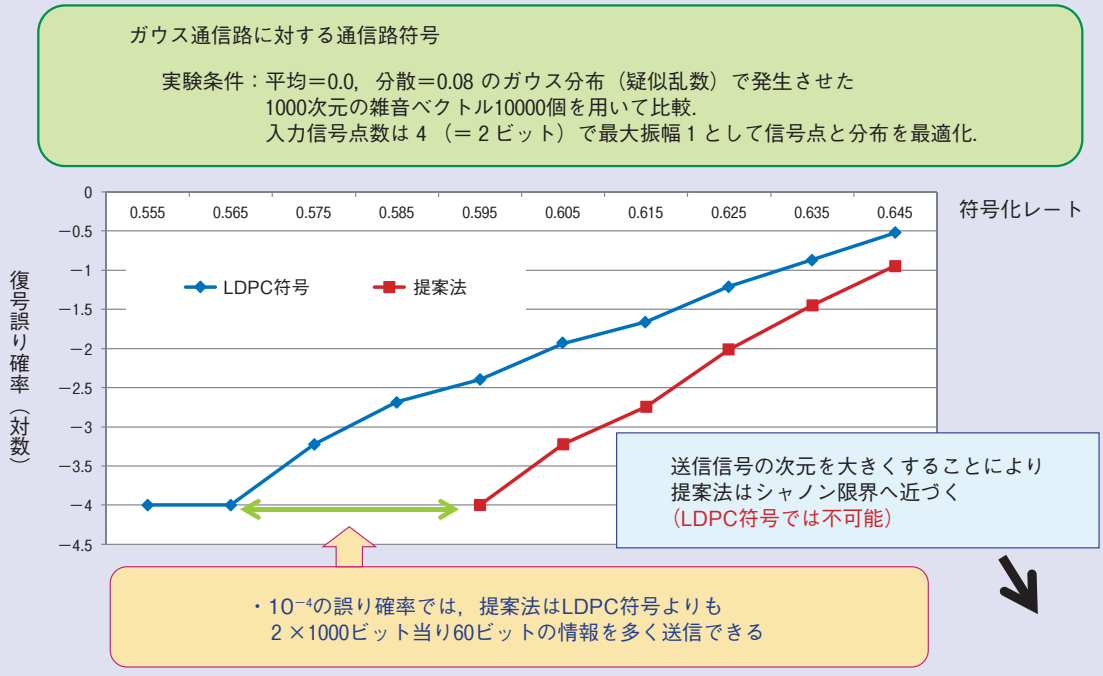
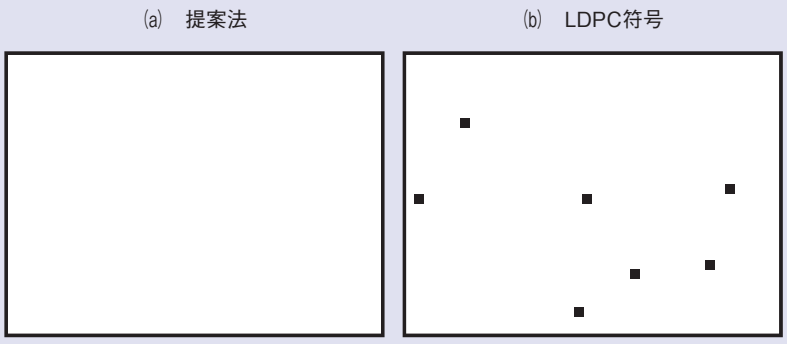


図5　シミュレーション実験



通信効率 0.6 の送受信を1200回繰り返したものを可視化.
黒点は 1 度の送受信（1200ビット）で 1 ビット以上の誤差があることを意味する.

図 6　通信効率をそろえて比較

図 7　圧縮画像（JPEG）を符号化して比較

たうえで符号化しています．このため 1 カ所でも復号誤りが起こればファイルが破壊され，画像の大部分は再生不可能となります．符号化レート0.5では，提案法も従来法のLDPC符号も正しく復号が行われていますが，符号化レート0.6のLDPC符号では，復号誤りが発生してファイルが破壊されていることが観測できました．このことからLDPC符号による符号化の限界は0.5と0.6の間にあり，提案法はそれを超える性能を持つことが確認できます．

今後の展開

今回実現した技術について，今後，実装のための周辺技術の確立を進め，実環境においてより高速な通信を実現するための要素技術として，本技術の応用の検討を進めます．

■参考文献

(1) J. Muramatsu and S. Miyake: “Concept of CoCoNuTS,” Proc. of AEW10 , p.4, Boppard, Germany, June 2017.
(2) J. Muramatsu: “Channel coding and lossy source coding using a generator of constrained random numbers,” IEEE Transactions on Information Theory, Vol.60, No.5, pp.2667-2686, 2014.
(3) J. Muramatsu and S. Miyake: “Channel code using constrained-random-number generator revisited,” IEEE Transactions on Information Theory, Vol.65, No.1, pp.500-508, 2019.

村松　純

車に例えると，情報理論は通信や情報処理の「エンジン」の研究になります．普段見ることはありませんが，エンジンがなければ車は走りません．当研究所では新しい「エンジン」の原理を生み出すことをめざしています．

◆問い合わせ先
NTTコミュニケーション科学基礎研究所
協創情報研究部 知能創発環境研究グループ
TEL　0774-93-5020
FAX　0774-93-5026
E-mail　cs-liaison-ml@hco.ntt.co.jp

主役登場

心を通わせて話せる対話ロボットをめざして

成松　宏美
NTTコミュニケーション科学基礎研究所
研究主任

子どものころ，お人形とごっこ遊びをしながら「見た目は人と同じなのにどうしておしゃべりはできないんだろう」と素朴な疑問を抱いたのを覚えています．お人形と物理的に「握手」をすることはできても，「言葉」で気持ちを伝え合うことはできませんでした．「人のように心を通わせておしゃべりできるロボットをつくりたい」これが私の対話研究をめざすきっかけとなりました．

人と心を通わせて話すロボットをつくるために，私は，人がどのように相手の発話を理解し，話しているのかを知り，ロボットの理解を1つひとつ積み重ねていく必要があると考えました．しかしながら，ロボットと人とがおしゃべりすることを目的とした対話研究の領域においては，たとえ短い間のやり取りであっても幅広い話題に対応できることが重要視されており，１つひとつ言葉のやり取りを丁寧に積み重ねて深い理解につなげるようなアプローチは，見方によっては古い考えとされることもありました．なぜなら，近年の対話研究では，ディープラーニングの発展とデータ収集の容易さゆえに，細かい理解によるルールベースの手法から，データ駆動型の手法へと主流が移ってきていたためです．一方で，私の周りの研究者たちは，人の発話を理解するために，人どうしの対話を分析し，地道に理解を積み上げることに対し，前向きに支持してくれました．それは，NTTコミュニケーション科学基礎研究所が，現在主流の一問一答ベースの雑談ロボットをいち早く実現し，次の課題を見据えることができていたことや，対話ロボットをつくるうえで，過去の先輩たちがつくった固有表現抽出や，述語項構造解析，焦点語抽出などの，言葉や文を解釈するのに重要な技術が有効であったからだと思います．

こうした対話ロボット実現に向けた多くの知見がある環境で，今回つくり出したのが，ユーザの体験した出来事（5W1H+感想）を文脈として理解し応答する雑談対話ロボットです．ロボットが，人の発話から5W1H+感想に該当する項目をフレーズとして理解し，それと同じような体験を根拠に共感を提示することで，人とロボットが心を通わせるところに一歩近づくことができました．主流に逆行する中で，周囲の支援を受けながら取り組んだフレーズ理解とチームが培ってきた対話の戦略のノウハウを組み合わせることにより，成し遂げることができました．

人と心を通わせて話すロボットは，家庭をはじめとした，人々の社会生活のさまざまなシーンにおいて，話し相手や悩みを相談できる良きパートナーとして，人々の暮らしに浸透していくと考えています．人に話すのは気が引けるという状況においては，人よりも気楽に相談できる相手になるかもしれません．人と心を通わせて話すロボットをめざすことは，目に見えない人の心を理解するという困難がある一方で，新しい発見・新しい技術につながる可能性が大きいと思っています．人のように思いやる心を持つロボットを夢に抱きながら，多くの人の良きパートナーになれるロボットの実現に向けて，引き続き検討を進めていきたいと思います．