

OS-3

質感と感性

Shitsukan and Kansei

坂本 真樹

Maki Sakamoto

電気通信大学

The University of Electro- Communications.
maki.sakamoto@uec.ac.jp

渡邊 淳司

Junji Watanabe

NTT コミュニケーション科学基礎研究所

NTT Communication Science Laboratories.
watanabe.junji@lab.ntt.co.jp

Keywords: shitsukan, texture, kansei, multimodal, perceptions.

1. はじめに

質感の認知は多感覚の情報処理を必要とする一方、感覚ごとに研究の進捗と方法論は異なっている。視覚の質感研究は研究が進んでいるとされるが、人の目に見えるレベルでの視覚質感認知メカニズムの解明、人工知能がどこまで人の視覚による質感認知メカニズムを再現可能なのかは未確認である。視覚に比べて原初的な感覚とされてきた触覚は、人工知能に触感を呈示認識、再生させる技術開発が待たれる。複雑な信号変調により生じる聴覚の質感認知、視触覚や嗅覚と切り離せない味覚の質感認知、高次感性処理の関係の解明も課題が多い。各感覚入力を通して得られる質感が複雑な質感概念ネットワークを構成している脳の全体像を解明するための方法論として、近年発展が著しいディープラーニングを含む統計的機械学習があるが、画像など視覚を対象とした研究が主となっている。本OSは、独立に研究が行われることの多いそれぞれの感覚の質感研究を、一つのセッションの中でその進捗や方法論を共有する機会とするという趣旨で設立した。人の五感と好き嫌いなどの価値判断に関わる理工系の研究（画像処理、触覚工学、音響学、機械学習、感性工学、言語処理）を中心としながら、知覚心理物理研究、脳神経科学といった心理・生理系の研究者との連携も図り、質感情報処理の科学的な理解と新たな人工知能研究の可能性について検討する場となることを期待して発表募集を行った。2018年度で3年目を迎えたが、一人15分の発表時間を確保するのが大変なほど応募があり、毎年2セッション設け、この分野に関連して最先端の研究が行われている招待講演者を毎年お招きし、全セッションの締めくくりには、共同オーガナイザの渡邊淳司とともに、分野を超えた議論を行う特別企画を行った。

毎年冒頭に5分ほど、オーガナイザの坂本真樹がセッションの趣旨説明を行った。2010～14年度科学研究費補助金新学術領域研究「質感脳情報学」に続き、2015～19年度科学研究費補助金新学術領域研究「多元質感知」が開始したことを契機に本セッションを提案したことなど背景を説明した。質感を生み出す情報は複雑な高次元情報として感覚入力に埋め込まれており、それを読み解く人の能力の解明は、人間の感覚情報処理の科学的理解のみならず、情報工学技術の発展にとっても不可欠な課題であり、人工知能の発展に寄与しているディープラーニングとも深いつながりがあることについて述べた。

2. 2016年度

記念すべき1回目OSの招待講演者には、東北大学岡谷貴之氏にご登壇いただき、「深層学習による材質・質感の画像認識」として、画像に写る物の材質やその表面



図1 2016年度の招待講演者岡谷貴之氏ご講演の様子

の質感を推定する問題を、畳込みニューラルネットワークによって解決しようとする研究について講演いただいた。例えば、材質カテゴリー認識と物体カテゴリー認識との関係、特に、物体認識から材質認識への転移学習の効果や、人の知覚との類似・相違点などについて紹介された。

一般発表では、電気通信大学の下田 和氏らが「質感画像の弱教師領域分割とその結果に基づく質感的部分的変換」と題する発表を行った。下田氏らは、Convolutional Neural Network (CNN) に基づく画像認識手法を応用して、質感画像の弱教師領域分割と、その結果に基づく画像的部分的な質感の変更に関する研究を行っており、次の2点について報告を行った。(1) CNNによる画像分類における画像中の各領域の認識結果に対する寄与を、バックプロパゲーションを応用した可視化手法をもとに分析し、それを応用した領域分割法を実現した。(2) Deep Style Transfer 手法を利用し、(1)の領域分割結果にその結果に基づく質感部分の変更を行い、その領域が変更後の質感カテゴリーの領域として(1)で認識されることを確認した。

NTTの米家 惇氏らによる「眼球運動および瞳孔径計測に基づく音の快不快の推定」は、音に対して感じられる快・不快の程度を、眼球計測に基づいて推定する手法を提案している。具体的には、協和音・不協和音を提示した際の眼球計測を行い、microsaccade (無意識下で生じる微小な跳躍性眼球運動) および縮瞳・散瞳を検出し、制御論モデルなどに基づく特徴量を抽出した。

NTTのHo Hsin-Ni氏は「Thermal modeling for material identification」と題して、触覚の質感の認知モデルに関する研究報告を行った。人間が物体に触れる際、皮膚温度変化は物体の材質に関する情報を有している。そのため、物体接触時の皮膚温度反応モデルは物体認識を行ううえで有用であるとしている。今回の発表では、構築したモデルの有効性について、物理的な皮膚温度変化との比較により報告している。

東京工業大学の永原 宙氏ら(共同オーガナイザ渡邊と共著)は「カテゴリー分類とMDS(多次元尺度構成法)に基づく知覚特性に対応した刺激空間構成法に関する考察」と題する発表を行った。多数の刺激の関係性を人間の知覚特性を反映した形で把握するためには、すべての刺激に対して複数項目で定量評価を行う必要があり、刺激の数が増加するにつれてその評価回数は莫大なものになってしまう。そこで永原氏らの研究では、被験者が多数の刺激を「似ている」-「似ていない」を基準に自由にグループ分けを行うだけの簡便な刺激空間構成法を提案している。

電気通信大学の鈴木航平氏ら(オーガナイザ坂本と共著)による「味や食感を表すオノマトペ感性評価システムの構築」は、本OSでは数少ない、味覚分野に関する研究報告であった。触覚・視覚・聴覚・嗅覚・味覚といっ

た五感すべてを通した経験が直感的な一言に込められるオノマトペが表す情報の数量化は、手触りなどを表すオノマトペが表す情報の数量化に比べて困難である。本研究では、オノマトペを構成する音韻と12個の味や食感、好き嫌いなどの感性に関わる尺度ごとの印象の音象徴的結び付きを利用し、食べたり飲んだりしたときに用いられるオノマトペの印象を数量化するシステムについて報告した。

東京大学の松田英子氏らによる「小学生における共感覚的傾向と認知モデルによる考察」は、共感覚という、ある感覚刺激に対し異なる種類の感覚が同時に感じられる認知現象からの報告であった。数字に対して色・擬人的な知覚を感じるタイプの共感覚について調査を行い、小学校4~6年生に対して年齢による変化を解析した結果、数字に対する擬人的な知覚は児童に広く見られること、学年が上がるにつれてその度合いが減少することがわかったとしている。

大阪大学の奥谷 凧氏らは「視覚的形狀操作において実物体表面質感を毀損しないステレオ投影の基礎検討」と題する発表を行った。プロジェクションマッピングにより、実物体を視覚的に形状操作するステレオ投影技術は、製品デザインプロセスを効率化することが期待できる。その一方、プロジェクタの空間解像度およびダイナミックレンジには限界があり、シボなどの表面質感の高品位な再現は困難とされる。そこで奥谷氏らは、投影対象の表面質感を利用するアプローチの研究を進めた結果得られた、視覚的形狀変形後も表面質感が保たれているように知覚されるという知見を報告した。

同じく大阪大学の青山宗太氏らは「重畳投影パターンの手に対する相対速度変調による力覚の操作」と題する報告を行った。青山氏らは、感覚間相互作用を利用して視覚刺激によって力覚を操作する研究を行っている。具体的には、平行移動するユーザの手指にドットパターンをその移動速度とは異なる速度で動かしつつプロジェクション投影して、ドットパターンの手に対する相対速度を変調する。これによって、擬似的な力覚を生じさせることができるかどうかを調査し、投影パターンの移動速度が、知覚する力の大きさに影響を与えることについて報告している。

以上のように、2016年度は多様な方法論を用いて嗅覚以外のすべての五感を通した情報処理に関する発表をいただくことができた。

3. 2017年度

2017年度は、画像分野同様、深層学習による進化が目覚ましい音声分野の研究で著名な名古屋大学の戸田智基氏を招待講演者にお招きし、「音声の声質を変換する技術とその応用」に関する講演をいただいた。音声には、話し手が何を話しているかという言語情報のみでなく、

誰がどのように話しているかという情報など、多種多様な情報が埋め込まれている。その中には物理的制約により決定されるものも数多く存在し、音声コミュニケーションにおける障壁を生む要因となるということで、物理的制約を超えた情報制御を可能とする声質変換技術について解説いただいた。さらにその応用例として、音声生成機能を拡張する技術についてデモ付きで聴衆の興味を引く紹介をしていただいた。

一般発表では、同じく音声分野から電気通信大学の中鹿 亘氏が「クラスタ適応制限ボルツマンマシンを用いた話者クラスタリングと声質変換への応用」と題する発表を行った。従来は、音声信号から自動的に潜在的な音韻情報と話者情報へ分離する適応制限ボルツマンマシンを用いて、入力話者音声から推定される音韻情報を保持したまま、話者情報のみを切り替えて、目標話者音声へ変換する声質変換手法を提案してきたが、中鹿氏らは、さらに話者情報を自動的にクラスタリングする拡張モデルとして、クラスタ適応制限ボルツマンマシンを提案し、声質変換タスクへ応用する手法について報告した。

この年は、さらに聴覚系の研究発表が続き、NTTの上村卓也氏が「聴覚系モデルを用いた音のテキスト変換」と題して、ある音の内容は保持したままテキストのみを変換する手法を提案した。過去に提案された聴覚系を模擬したモデルによる音のテキスト合成と、畳込みニューラルネットワークによる画像のテキスト変換を組み合わせた手法である。異なる2音から計算したモデルの出力波形をもとに、それぞれの特徴を受け継いだ新たな音を合成するというものである。例えば、音声や音楽に適用することで水中でしゃべったような音声したり、鳥の鳴き声のような音楽を合成している。

まだ人工知能での研究が少ない触覚分野からは、NTTの横坂拓巳氏が「手の動きと結び付いた触質感認知の研究」と題する発表を行った。我々は身の回りの物体の触質感を感じ取るために、手や指を動かして能動的にその物体を探索している。手の動きと触質感認知の結び付きへの注目が高まっているが、その体系的な関係はまだまだ

明らかでない。そこで横坂氏らの発表では、どのような手の動きが、粗さや硬さといった物性的質感、布らしさや木らしさといった素材的質感、心地良さのような感性的質感と結び付いているのかについて調べた心理物理実験について報告を行った。

2016年度に続き、視覚系の研究発表は多かった。電気通信大学の権 眞煥氏ら(坂本・共著)の「DCNNを用いた画像の質感認知—音象徴性からのアプローチ—」と題する研究発表は、いわゆる質感を直感的に表現する際に用いられる「つるつる」、「さらさら」といったオノマトペに深層学習を適用したものであった。人は画像を見た際に膨大な特徴を瞬時に判断し、容易にかつ正確に質感を認知することが可能で、日本人は直感的にオノマトペで質感を表現する。特に、オノマトペを構成する音韻に質感印象が結び付くことから、画像から音韻を確率的に出力する畳込みニューラルネットワークの応用について報告した。

さらに、電気通信大学の下田 和氏らは「画像スタイル変換と Web 画像を用いた画像の任意質感生成」について発表した。下田氏は Neural Style Transfer アルゴリズムを実世界画像の質感の転写に応用し、コンテンツ画像と質感単語による画像内物体の任意質感変換を目指している。質感単語から質感の視覚的表現を自動構築するため DCNN 中間層の出力と画像のスタイル認識のために考案した特徴量 Neural Style Vector を用いて Web 画像から有用なスタイル表現を自動選択するシステムについて報告した。

同じく電気通信大学の丹野良介氏は「複数スタイルの融合と部分的適用を可能とする multi-style feed-forward network の提案」と題する発表を行った。CNN によるスタイル変換はバックプロパゲーションを利用してため時間が掛かることが問題になっていた。それに対して feed forward で変換するネットワークを実現する方法が提案されている。丹野氏らの発表では、そのネットワークに conditional な入力を追加して、複数スタイルの混合、部分的な適用を可能にする network を提案している。

大阪大学の奥谷 凪氏は、「ステレオ投影による視覚的変形で知覚される実物体表面質感の調査」と題する発表を行った。プロジェクションマッピングにより、実物体を視覚的に形状操作するステレオ投影技術は、製品のデザインを効率良く行うことが期待できる。一方で、プロジェクタの空間解像度およびダイナミックレンジには限界があり、細かな表面質感の高品位な再現は困難という課題がある。そこで奥谷氏は、投影対象の表面質感をそのまま利用するアプローチを提案し、視覚的形状変形後も保持されるように知覚される表面質感の条件を調査した。



図2 2017年度招待講演者戸田智基氏ご講演の様子

4. 2018 年度

2018 年度は、視覚系や聴覚系に比べて人工知能技術の適用が遅れている触覚系の研究について意欲的な試みを行っている名古屋工業大学の田中由浩先生に招待講演を依頼した。田中氏は「個人の触覚情報の多様性と活用」というタイトルの講演を行った。触覚は極めて身体性の高い情報であり、人工知能にデータを取得させるためのインタフェースが課題であるが、人間は外界との物理的接触を通じて、自身の身体の力学的変化、皮膚の変形や振動、温度変化をもとに触覚的情報を取得している。人間が知覚している触覚情報は、自身の皮膚の状態を含んでおり、さらには運動に応じてその情報は変化する。人間は運動を通じて触覚を得るが、同時に触覚を通じて運動を変化させる。このように両者には双方向の関係があるため、人工知能と人の相互作用が不可避となる。さらに厄介なことに、触覚は、皮膚や運動といった力学的特性においても個人差があり、さらに感度や認知の個人差もある。極めて主観的であり多様な感覚世界、個性を触覚情報は表現していると考えられる。しかし、田中氏によれば、このような多様性を捉え、活用するために、従来の官能評価や心理物理学的手法に加えて機械学習は有効であり、他者理解や人の能力を活用したり人に優しいデザインにつながるとしている。主観的な触覚の情報化と活用は始まったばかりであるが、田中氏の講演は、個人の触覚を情報化する取り組みや、それを活用した人の知覚研究や応用研究を紹介し、機械学習活用の可能性と発展について示唆を与えるものであった。

一般発表は、残念ながら触覚分野の研究はなかったが、3年連続本 OS で発表してくださっている研究チームも多く、質が高まっていることを実感できた。

3年連続で本 OS に貢献していただいている NTT の上村卓也氏は、「多様な質感認識の情報処理に用いられる画像特徴を統一的に説明するためのニューラルネットワークモデルの検討」と題する研究発表を行った。自

然な視覚刺激には光沢・半透明性・物体の素材などの豊富な質感が含まれる。一般に物を見る際には数多くの視覚特徴が処理されるため多様な質感属性を統一的に説明することは難しいという課題に対し、質感知覚に用いられる視覚特徴を説明するために、質感弁別課題の実験データを分析したことについての報告だった。実験参加者の回答が刺激の画像特徴に基づいていると仮定し、画像特徴は画像分類に最適化された深層ニューラルネットワーク (DNN) によって計算した。画像分類に最適化された DNN では、深い層ほど複雑で抽象的な特徴が表現されるといわれるが、実験参加者の反応を最も良く説明したのは中間の層の特徴であったという興味深い報告をしていた。この結果から、質感知覚には比較的高次の特徴が重要であることが示唆され、質感の属性ごとに用いられる特徴が異なることもわかったとしている。DNN によって質感知覚に用いられる視覚特徴を効果的に説明できることが示されたことは意義が大きい。

稲住朋彦氏ら (坂本・共著) が、室内空間における雰囲気考慮した素材提案システムの構築に関する研究報告を行った。インテリアを対象とした研究はさまざまな分野で行われており、室内空間のシミュレーションに関する研究や最適化に関する研究、壁や床に使用される素材の印象評価に関する研究など多岐にわたる。室内空間における全体的な雰囲気や印象は、作業や居住をする人々の感情や行動に影響を与えるため、ユーザが要求する雰囲気に沿った手軽で直感的なインテリア提案手法が求められる。そこで、この研究ではインテリアにおけるデザインとしての役割に着目し、ユーザが要求する室内空間の雰囲気に合わせた壁や床などの素材を提案するシステムを構築したというもので、質感分野の応用研究である。室内空間の全体的な雰囲気を表現する言葉として、直感的かつ微細なニュアンスを表現できるオノマトペを入力とし、ユーザによる直感的な要望に基づきデザイン提案が行える可能性を示した。

電気通信大学の柳井啓司研究室からも、3年連続で本 OS での発表をいただいた。2018 年度は、「画像内容を考慮した質感表現に基づく画像変換」と題して杉山 優氏に発表いただいた。近年、計算機の許容する計算量が増えたことにより、深層学習による画像の生成や変換が注目されている。なかでも画像変換の技術として代表的なものにスタイル変換と呼ばれる技術がある。しかしこれはスタイル画像とコンテンツ画像を、時間をかけて最適化しなければ変換が行えない。改良手法の高速スタイル変換では、あらかじめ学習したスタイル画像が決まっていればどのような画像も似たような変換になってしまうという課題があった。そこで本研究では、言葉による画像変換を実装すると同時にコンテンツ画像の物体認識を行った。コンテンツがどういった画像かということに対して適切な変換を行うことによって、コンテンツ画像ごとに異なる適切な変換を言葉によって指定された方向性



図 3 2018 年度招待講演者田中由浩氏ご講演の様子

に画像変換することが可能となったとしている。

同じく柳井研究室からは「深層学習による質感文字生成」と題して、下田 和氏らが文字のフォントの自動生成、自動変換のタスクに取り組んでいる。従来のフォント生成のタスクでは文字をいくつかのストロークから成り立つものとし、モデル化を行い作成する手法が取られてきた。それに対し本研究では深層学習により、画像中のフォント画像ないしパターン画像からストロークに相当する特徴を自動で抽出し、変換元のフォントから任意のデザインパターンへの変換に挑戦している。この仕組みにより手書き文字のような個人ごとのオリジナルフォントの作成が容易にできるようになる。実験ではケチャップ文字をはじめとしたユニークな質感パターン画像セットを作成し、深層学習のクロスドメイン学習による手法と **Neural Style Transfer** の手法とを組み合わせ、生成結果の可読性を改善したとしている。

電気通信大学の中鹿 亘研究室も2年連続で発表していただいた。2018年度は「RBMを用いた楽器音基底と演奏情報への分離による多重音解析」と題して、荒川賢也氏が研究発表を行った。自動作曲など発展的な音楽研究では、膨大な楽譜データが必要とされる。そのために波形信号から楽譜を自動生成することも重要な課題である。楽器音分離においては **non-negative matrix factorization (NMF)** がよく用いられているが、精度には向上の余地がある。そこで本発表では、生成モデルの **restricted Boltzmann machine (RBM)** を用いた楽器音分離を提案している。二つのアルゴリズムの大きな違いとして、**NMF** が決定的に解を求めるのに対して **RBM** は確率的に解を求めるという点、**NMF** は演奏情報行列が連続値で出力されるのに対して **RBM** は楽譜表現に適したバイナリ値で出力される点あげられる。例えば、生演奏から楽譜を生成する場合、**NMF** よりも確率分布で議論を行う **RBM** のほうが入力のおわずかな差異に柔軟に対応できると考えられる。本研究ではそのような入力への柔軟性および出力形式の違いから **RBM** および **RBM** を拡張した **sparse-RBM** を用いて音楽の波形信号を楽器音基底行列とその演奏情報行列に分離する実験を行った。結果として提案手法は **NMF** と比較してわずかに良い結果を示したとしている。

5. おわりに

以上、3年間の本OSの研究発表を紹介してきたが、質感研究が特定のモダリティーに閉じた研究ではなく、複数のモダリティーにまたがる多様性があること、人工知能技術が質感研究においていかに重要であり、かつ質感研究が人工知能の新たな課題を浮彫りにするものであるかを感じていただけたのではないかとと思われる。

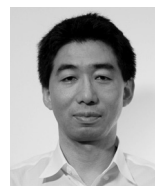
OSの統括の際には、質感・感性に関する研究者が人工知能研究に期待すること（人工知能が質感・感性研究に貢献できること）、そして人工知能分野の研究者が質感・感性の研究に期待すること（質感・感性研究が人工知能に貢献できること）について議論した。本OSは、人工知能学会員以外の方も多く集まり、特に産業界からの参加も多かったようである。本OSらしく、生体データの扱いなど、データセットの問題や方法論について幅広い視点から議論が展開された。本OSは2018年度で3年目となったが、「質感」というキーワードの人工知能分野における重要性を認識していただけたのではないかと期待される。

著者紹介



坂本 真樹 (正会員)

1998年3月東京大学大学院総合文化研究科言語情報科学専攻博士課程修了。東京大学大学院総合文化研究科言語情報科学専攻助手、電気通信大学電気通信学部准教授を経て、2015年4月より同大学院情報理工学研究所教授。情報処理学会、日本バーチャルリアリティ学会、感性工学会、日本認知科学会などの各会員。



渡邊 淳司

NTTコミュニケーション科学基礎研究所人間情報研究部主任研究員（特別研究員）、東京工業大学工学院特任准教授兼任。博士（情報理工学）。人間の触覚の知覚メカニズム、感覚を表現する言葉の研究を行う。人間の知覚特性を利用したインタフェース技術を開発、展示公開するなかで、人間の感覚と環境との関係性を理論と応用の両面から研究している。近年は、学会活動だけでなく、出版活動や、科学館や芸術祭において数多くの展示を行う。主著に『情報を生み出す触覚の知性』DOJIN SENSYO (2014)（毎日出版文化賞（自然科学部門）受賞）がある。