

2019年4月1日から2020年3月24日の期間で、英国サリー州にある University of Surrey の Centre for Vision, Speech and Signal Processing (CVSSP) に滞在し、研究活動を行った。CVSSP は、機械学習を用いたマルチメディア向けの音響信号処理、画像信号処理、生体信号処理などを中心に精力的な研究活動を行っている研究所で、英国放送協会 (BBC) などとも共同研究を行っている。また、研究所内の研究者および博士課程の学生は非常に多様な国から来ており、様々な文化が混在した多国籍な点も、CVSSP の特長である。さらに、CVSSP には様々な国から多くの客員研究員および短期留学生も滞在している。私が滞在した研究室は、CVSSP の中で音響信号処理研究の中心となっている研究室で、Wenwu Wang 教授が主催している。

University of Surrey の CVSSP で私は、音響信号処理に関する基礎理論の研究を中心に行った。具体的には、(1) 音響信号に代表される概周期信号を構成する正弦波のパラメータ推定に関する研究、(2) 複数の音響信号が混合された観測信号から個々の信号を推定する信号分離に関する研究、および (3) 音響信号に含まれるパラメータの加工に関する研究を行った。

(1) 正弦波のパラメータ推定に関する研究

本研究では、概周期信号を構成する正弦波のパラメータの1つである周波数を、少ない演算量で正確に推定する技術の開発に従事した。正弦波の周波数は、音響信号などの概周期信号を特徴付ける非常に重要な要素で、その推定に関する研究は古くから行われているが、未だに決定的な方法は確立していない。私が取り組んだ研究は、時間領域で観測した信号を周波数領域と呼ばれる状態に変換する離散フーリエ変換 (DFT) を利用した周波数推定法の1つである interpolated DFT に関するものである。DFT を用いると、時間領域では周波数が分かりづらい信号でも、スペクトルとよばれる波形のピークの位置で周波数を知ることができる。一方、DFT はコンピュータを使用した信号の変換方法であり、得られるスペクトルは離散的な値となるため、周波数

も離散的に得られる。interpolated DFT では、スペクトルのピークおよびその周辺のスペクトルの値を利用して、スペクトルを連続値で表した場合に得られる真のピーク位置を推定する。interpolated DFT は 1990 年代から研究が行われているが、2000 年代後半になり正弦波を構成する要素である単一の複素正弦波における研究が飛躍的に進んだ。一方、複数の複素正弦波によって構成される一般的な概周期信号に関する研究は、現在に至るまであまり進展していない。

複数の複素正弦波から成る信号の DFT スペクトルから周波数を推定する場合、各複素正弦波に由来するサイドローブと呼ばれる成分が互いに干渉するため、正確な周波数の推定が困難となる。一方、サイドローブは窓関数と呼ばれる重みを時間領域で信号に乗算することで低減できることが古くから知られており、実際に多くの信号処理技術で窓関数が利用されている。そこで、本研究では窓関数を interpolated DFT にも導入するために、窓関数を乗算した複素正弦波の DFT スペクトルについて、その数式モデルの正確な記述を行った。そして、窓関数を用いない従来の interpolated DFT アルゴリズムを、この数式モデルによって窓関数を用いた場合へ拡張することに成功した。シミュレーションによって、この方法は複数の複素正弦波が含まれる信号に対しても有効であることが示された。この研究成果は、信号処理研究の分野では世界最大の規模を誇る国際会議である 45th International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2020) に投稿、受理された。

また、これらの研究に関連して、従来の interpolated DFT アルゴリズムが持つ様々な欠点を改善することにも成功しつつある。これらについては、現在も Wenwu Wang 教授と共同で研究を継続している。

(2) 信号分離に関する研究

本研究は、非負値行列因子分解 (NMF) と呼ばれる技術の音響信号分離への応用に関するものである。NMF は、非負値要素から成る 1 つの行列を同じく非負値要素から成る 2 つの行列の積に分解する技術である。近年、音響信号を時間周波数領域に変換して得られるスペクトログラムと呼ばれるものに NMF を適用すると、複数の音響信号が混合された信号の場合に、条件によって個々の音響信号のスペクトログラムに分解することができることが報告されている。本研究では、NMF による音響信号分離が可能となる条件について、従来よりも幅広い音響信号に適用可能な信号モデルの構築を行った。

従来の NMF を用いた音響信号分離では、NMF に必要となる最適化問題の解きやすさからガウス分布などの分布の裾が比較的急峻に減衰する確率分布が

用いられてきた。一方、実際の音響信号のスペクトログラムの確率分布を観察するとガウス分布などよりも裾が重い分布となっていることが多く、これが原因となり NMF による信号分離性能が限定的となっている可能性が示唆されていた。そこで、本研究では裾が重い分布の 1 つである複素ラプラス分布を用いて信号モデルを構築し、さらに上界最小化 (MM) アルゴリズムを応用して収束を保証した最適化アルゴリズムを導出した。この研究成果は、信号処理の分野では欧州最大の規模を誇る国際会議である 27th European Signal Processing Conference (EUSIPCO 2019) に投稿、受理されて発表を行い、世界中の専門家らとディスカッションおよび情報交換を行なった。

また、構築した信号モデルをさらに幅広い音響信号に適用可能とするために、さらなる信号モデルの一般化を行った。ここでは、ベッセル関数分布と呼ばれる確率分布を用いて、従来から用いられているガウス分布およびラプラス分布を包含した信号モデルを構築し、それにもとづく評価関数を MM 法によって最適化することに成功した。この研究成果は、アジア地域の信号処理分野の研究者が中心となって開催される国際会議である Asia-Pacific Signal and Information Processing Association Annual Summit and Conference 2019 (APSIPA ASC 2019) に投稿、受理されて発表を行なった。

(3) 音響信号のパラメータ加工に関する研究

本研究は、音響信号の話速・音高変換を行うためのアルゴリズムである phase vocoder に関するものである。話速変換は音響信号が持つ音の高さを変化させることなく音響信号の再生速度のみを変化させる技術であり、また、音高変換は音響信号の再生速度を変化させずに音の高さのみを変化させる技術である。phase vocoder は、音響信号を時間周波数領域に変換した後にスペクトログラム上の位相成分を加工することで、音高や話速の変換を行う。

phase vocoder は複数の音響信号が混合した場合でも話速・音高変換を行うことができるが、得られる信号には phasiness と呼ばれる残響に似た雑音の発生や transient smearing と呼ばれる音の瞬間的な変化の鈍化といった音質の劣化が発生することが知られている。前述の「(1) 正弦波のパラメータ推定に関する研究」も、この phase vocoder における音質劣化の改善に向けた取り組みの一環であるが、ここでは phase vocoder の問題点をあらためて整理することで、次に取り組むべき課題の洗い出しを行った。その結果、transient smearing の軽減に注力するべきであるとの結論が得られた。本研究は、CVSSP の客員研究員で Wenwu Wang 教授の共同研究者でもある中国の四川大学の Shiyong Lan 博士と共同で実施し、本研究成果を含めた phase vocoder についての講演を四川大学において行なった。

以上のように、本在外研究では多くの研究成果が得られ、そのうちの一部は現在も共同研究として継続中である。また、CVSSP には世界中から様々な研究者が訪れるため、信号処理に関する様々な分野において最先端で活躍する多くの研究者と交流を持つことができ、非常に有意義な滞在であった。これらの成果を、自身の研究および研究室のみならず、学科や学部、大学にも何らかの形で還元できるよう努める所存である。