様　式　Ｃ－４１

科学研究費助成事業　　研究成果報告書

令和　３ 年　 ６ 月 １０ 日現在

機関番号：８２４０１
研究種目：奨励研究
研究期間：2020～2020
課題番号：２０Ｈ０１１５９

研究課題名　　話者・言語特徴の分離表現に基づく音声強調・認識の統合

研究代表者
関口　航平（Sekiguchi, Kouhei）

国立研究開発法人理化学研究所・革新知能統合研究センター・テクニカルスタッフ１

交付決定額（研究期間全体）：（直接経費）　　480,000円

研究成果の概要：音声認識を用いるシステムにおいて、マイクと話者の距離が離れている場合には周囲の雑音や残響などもマイクでの観測音に含まれてしまうために音声認識が困難になる問題がある。そのため、観測音から音声のみを取り出す研究が盛んにおこなわれている。本研究では、使用する環境を特定しない汎用的な音声強調手法に深層学習を用いた音声の生成モデルを統合した従来手法に着目し、音声が時不変な話者情報と時変な言語情報に依存するという性質を考慮した音声の生成モデルを用いることで、音声強調精度のさらなる改善を図った。

研究成果の学術的意義や社会的意義
音声認識は、スマートフォンなどのように話者とマイクの距離が近い場合には、現在すでに高い認識率を達成しているものの、話者とマイクの距離が離れている場合には周囲の影響により認識率は大幅に低下してしまう。このような状況における認識率を改善することができれば、スマートスピーカなどを快適に利用できるようになったり、聴覚障害者の日常生活を補助するようなデバイスを実現することが可能になったりすると考えられるため、音声強調は重要な研究テーマである。

研究分野：統計的音響信号処理

キーワード：音声強調　音源分離　音声認識

１．研究の目的

スマートスピーカやロボットの音声対話においては話者がマイクから離れた位置にいる場合があり、このような状況ではマイクの観測音に周囲の雑音や残響が含まれてしまい音声認識の精度が低下してしまう問題がある。そのため、マイクの観測音から注目したい話者の音声のみを取り出す音声強調手法が必要となる。本研究では、様々な環境で汎用的に用いることのできる手法を開発することを目的とする。

２．研究成果

これまでの研究で、環境の事前情報を用いず、クリーンな音声の生成過程のみをニューラルネットワークで学習し、それを用いて多チャネル観測音の生成モデルを定式化し、その逆問題を解くことによって各音源信号を推定するという手法を提案している。本研究では、この手法を拡張し、音声が話者情報と言語情報から生成されるという生成モデルを用いることで、音声強調問題を話者情報と言語情報を観測音から推定する問題とみなす（図 1)。この言語情報を入

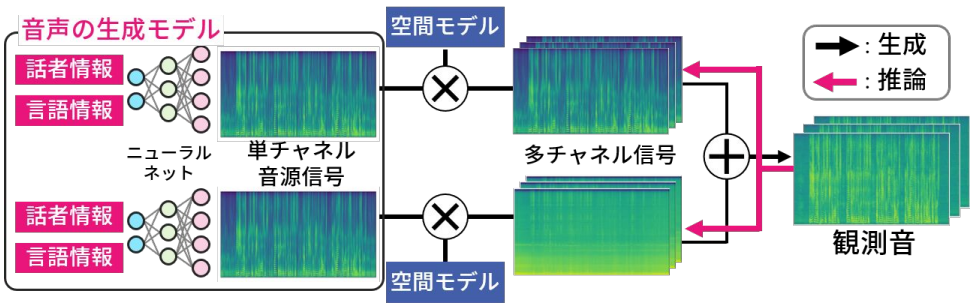

図 1　観測音の生成モデル

力とする音声認識器を学習することができれば、話者に依存しないために少ないデータで頑健な音声認識器が学習できる可能性がある。また、音声強調時に話者が既知である場合は、言語情報のみを推定すればよいため解空間が狭まり性能が向上する可能性もある。そのため本研究ではまず音声の話者・言語情報への分離と、話者・言語情報からの音声の生成過程を学習した。具体的には、条件付き変分自己符号化器（CVAE）を用いて入力音声を低次元の潜在表現に変換し、潜在表現から入力の音声を復元するニューラルネットワークを学習する。低次元な潜在表現を時不変な部分と時変な部分で構成し、さらに様々な制約を課すことにより。それぞれを話者情報と言語情報に対応させることを試みた。学習した CVAE に対し、学習データ中の音声を入力した場合には元の音声を復元することができたが、言語情報を保持したまま話者情報のみを変えた場合に得られた音声は、多少話者性は変わっているものの目的の話者の音声とは大きく異なっていた。また、話者情報を保持したまま言語情報を変えた場合も、得られた音声は元と異なっているものの目的としていたものとは異なっていた。つまり、話者情報と言語情報の分離の精度が低いことが分かり、この点については大きく改善の余地がある。しかしながら、話者情報と言語情報に基づく生成モデルを統合した音声強調手法の評価では，多少の性能の改善が確認できた.

主な発表論文等

〔雑誌論文〕　計1件（うち査読付論文　1件／うち国際共著　0件／うちオープンアクセス　1件）

１．著者名 Kouhei Sekiguchi, Yoshiaki Bando, Aditya Arie Nugraha, Kazuyoshi Yoshii, Tatsuya Kawahara	４．巻 28
２．論文標題 Fast Multichannel Nonnegative Matrix Factorization With Directivity-Aware Jointly-Diagonalizable Spatial Covariance Matrices for Blind Source Separation	５．発行年 2020年
３．雑誌名 IEEE/ACM TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING	６．最初と最後の頁 2610-2625
掲載論文のDOI（デジタルオブジェクト識別子） 10.1109/TASLP.2020.3019181	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 －

〔学会発表〕　計2件（うち招待講演　0件／うち国際学会　1件）

１．発表者名 Yicheng Du, Kouhei Sekiguchi, Yoshiaki Bando, Aditya Arie Nugraha, Mathieu Fontaine, Kazuyoshi Yoshii, Tatsuya Kawahara
２．発表標題 Semi-supervised Multichannel Speech Separation Based on a Phone- and Speaker-Aware Deep Generative Model of Speech Spectrograms
３．学会等名 2020 28th European Signal Processing Conference (EUSIPCO)（国際学会）
４．発表年 2021年

１．発表者名 關口 航平，坂東 宜昭，ヌグラハ アディティヤ，フォンテーヌ マシュー，吉井 和佳
２．発表標題 ARMA-FastMNMFに基づく同時的ブラインド音源分離・残響除去
３．学会等名 日本音響学会 2021年春季研究発表会
４．発表年 2021年

〔図書〕　計0件

〔産業財産権〕

〔その他〕

－

研究組織（研究協力者）

氏名	ローマ字氏名