



TOHOKU
UNIVERSITY



WASEDA University
早稲田大学



筑波大学
University of Tsukuba



内閣府

**倒壊瓦礫内に進入し、声を聞き取ることにより、被災者を発見
内閣府タフ・ロボティクス・チャレンジによる索状ロボット「能動スコープカメラ」
聞き取り能力の飛躍的向上により、地震災害の救助を高度化**

内閣府総合科学技術・イノベーション会議が主導する革新的研究開発推進プログラム（ImPACT）タフ・ロボティクス・チャレンジ（プログラム・マネージャー：田所 諭）の一環として、東北大学田所諭（たどころさとし）教授、昆陽雅司（こんようまさし）准教授、早稲田大学奥乃博（おくのひろし）教授、京都大学糸山克寿（いとやまかつとし）助教、坂東宜昭（ばんどうよしあき）特別研究員、東京大学猿渡洋（さるわたりひろし）教授、筑波大学牧野昭二（まきのしょうじ）教授、国立情報学研究所小野順貴（おののぶたか）准教授のグループは、瓦礫内捜索用ヘビ型ロボット「能動スコープカメラ」に複数のマイクロホンを搭載し、音響信号処理に基づく音声強調技術によって、瓦礫奥深くの要救助者が発する声を聞き取るシステムの開発に成功しました。

ポイント

- 索状ロボット「能動スコープカメラ」に複数の超小型マイクロホンを搭載。
- 新たに開発した、音響信号処理に基づく以下の2つの音声強調技術によって、騒音にかき消された瓦礫内の要救助者の声を抽出・強調して聞き取ることが可能に。
- 新たな音声強調技術の1つ、VB-MRNMFによる方式では、リアルタイムに聞き取ることが可能。
- 他方の音声強調技術 IVA+ポストフィルタによる方式では、より鮮明な音声を得ることが可能。

<研究の背景と経緯>

熊本地震や阪神淡路大震災に代表される大規模地震災害では、倒壊した建物内に取り残された人の発見と救助が大きな課題です。要救助者が発する助けを求める声は発見の重要な手がかりですが、災害環境での各種騒音にかき消され、瓦礫の外からそれを聞き取ることはたいへん困難でした。索状（ヘビ型）ロボット「能動スコープカメラ」は瓦礫内に進入して要救助者を発見することを目的として開発されたロボットですが、ロボットの運動に伴って発生するノイズが、瓦礫内での声の聞き取りにおける大きな障害でした。

東北大学田所教授、昆陽准教授、早稲田大学奥乃教授、東京大学猿渡教授を中心とするグループは、音声聞き取り性能の向上を目的として、ImPACT タフ・ロボティクス・チャレンジの共同研究開発の一つのテーマとして、能動スコープカメラに搭載可能な音声抽出・強調技術の開発を進めてきました。熊本地震や阪神淡路大震災の被災木造家屋を模擬した評価試験フィールドにて実証試験を行い、一定の性能が確認されました。

なお、ImPACT タフ・ロボティクス・チャレンジで研究開発を進めるレスキューロボットは、1995年の阪神淡路大震災を契機に田所らのグループが提唱し、世界の研究を牽引してきた研究分野です。また、ロボットにマイクをつけ、混合音を聞き分けるロボット聴覚の研究は、2000年に奥乃らが世界で初めて提唱した日本発の研究分野であり、以降、同グループは世界の研究を牽引してきました。これらの研究開発は世界的に高く評価されており、田所教授が災害ロボットの研究で、奥乃教授がロボット聴覚の研究で、牧野教授がブラインド音源分離の研究で、IEEE Fellow（米国電子技術協会 IEEE の最高位メンバー資格）に昇格しています。また、国内でも、科学技術分野での文部科学大臣表彰科学技術賞（研究部門）を、奥乃グループが H25 年度にロボット聴覚の研究で、猿渡グループが H27 年度にブラインド音源分離の研究で受賞しています。

<研究成果の概要>

能動スコープカメラは、内視鏡の表面に分布型振動駆動のためのアクチュエータを搭載することによって、瓦礫や配管などの狭く深い場所への進入を可能にした索状（ヘビ型）ロボットです。このロボットは数 cm の瓦礫の隙間をぬって深さ数 m まで探査できる世界で唯一の性能を有しており、これまで、建設現場倒壊事故原因調査などに使用され、福島第一原発建屋内の調査にも使用が検討され、細い配管内部の点検に性能を発揮するなど、優れた実績を上げてきました（図1）。

今回、能動スコープカメラの構造を、柔軟チューブを用いた新しい設計に変更することによって、バス方式で接続された多数のマイクロホン（以下、マイク）を一定間隔で搭載し、マイクアレイを構成することが可能になりました（田所、昆陽担当）。能動スコープカメラを瓦礫内に進入させると、多数のマイクを瓦礫内に分布させることとなり、要救助者の音声を多数の位置で、位相情報も含めて取得することが可能です（図2）。



図1 能動スコープカメラ



図2 搭載した複数のマイク

このシステムの特性を活かした音声抽出・強調を行うために、新たに、VB-MRNMF による方式（奥乃、糸山、坂東担当）、および、IVA+ポストフィルタによる方式（猿渡、牧野、小野担当）を開発し、能動スコープカメラに搭載することに成功しました。前者は、リアルタイムで音声強調が可能な方式であり、主として瓦礫内で要救助者を検索する際に効果を発揮します。後者は、数十秒の事後解析の処理時間が必要ですが、より鮮明な音声を得ることができるため、主として要救助者からの状況や健康状態の聞き取りなどに効果を発揮します。

熊本地震の現地調査に基づき、倒壊木造家屋の内部を模擬した試験評価用瓦礫を製作し、本システムの実証実験を行ったところ、これまでと比較して飛躍的に高い聞き取り性能を確認することができました（図3、図4）。



図3 屋外フィールド評価のための瓦礫



図4 熊本地震の倒壊木造家屋の内部を模擬した瓦礫空間への能動スコープカメラの進入

<音声抽出・強調技術について>

索状ロボットに装着したマイクアレイ（複数のマイクを使用、配置は任意）から取得した音から音声を強調する上で3つの技術的課題があります。

1. 走行雑音（自己雑音と呼ぶ）の変化が非定常であるので、事前に自己雑音を録音し、そのテンプレートの事前学習を行っても、有効ではありません。
2. ロボットの移動に合わせてマイク配置が動的に変動するので、通常のマイクアレイ処理^{注1)}を適用しても、性能ができません。
3. 瓦礫で一部のマイクの音声の音量が小さくなり、かつ、音量変化がどのマイクで生じるか事前にはわからないので、すべてのマイクからの入力が信頼できません。

この3つの課題に対して、音響処理の方式は2つの相反する要求条件—処理性能と処理速度—のトレードオフに対応する必要があります。そこで、ImPACT タフ・ロボティクス・チャレンジでは2つの解決策を開発しました。

手法 A（オンライン処理）：遅延時間が少なく、ほぼリアルタイムで性能は許容範囲内。

手法 B（オフライン処理）：性能は優れるが、処理時間がデータ量の2乗に比例してかかる。

手法 A は、奥乃グループ（早稲田大学、京都大学）が、手法 B は、猿渡グループ（東京大学、筑波大学、国立情報学研究所）が開発しました。いずれの方法も、入力音のスペクトログラムの大部分が雑音であり、音声は少ないという「スペクトログラム上のスパース性」^{注2)}を活用したスパース信号処理に基づいた手法を使用しています。

手法 A は、スパース信号処理と各マイクの信頼度を推定して、音声強調を行います。本手法は、VB-MRNMF (Variational Bayesian multi-channel robust NMF (Nonnegative Matrix Factorization, 非負値行列因子分解)^{注3)})と呼んでいます。具体的には、次のような処理で音声強調を行います。

1. 低ランク・スパース分解に基づく事前知識不要のブラインド信号処理。
2. マイク配置に頑健な振幅スペクトログラム領域^{注4)}での信号処理。
3. 各マイクの音声の音量を信頼度として推定。

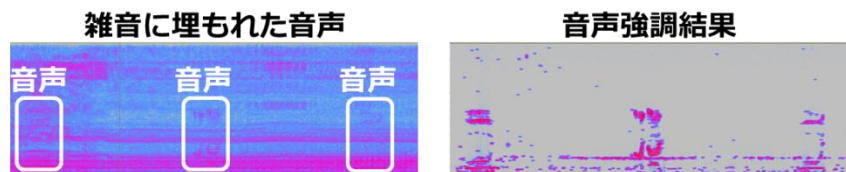


図5 入力音のスペクトログラムと音声強調結果のスペクトログラム

VB-MRNMF は、図5のように、雑音、とくに走行音などの自己雑音や外部雑音に埋もれた音声をノート PC 上でリアルタイムに強調することができます。また、障害物等で半数のマイクが覆われたり隠されたりしても、頑健に音声強調をすることができます。

手法 B は、IVA (独立ベクトル分析)^{注5)}と統計的ポストフィルタ処理で音声強調を行います。IVA は、手法 A とは別タイプの音源分離手法であり、音源モデルを「1 ランク構造を持つスパーズ時間・周波数スペクトログラム (NMF の単純な場合に相当)」、音の空間伝搬モデルを「線形時不変フィルタ^{注6)}」で表し、両者を一切のトレーニングシーケンスを使わず個別に自律推定します。これは特に、空中を伝搬する人間の声を表すのに適したモデルであり、比較的歪みの少ない分離音を出力することが出来ます。今回は被災現場における過酷な音環境に鑑み、この IVA をよりロバストにするため、以下2種類の拡張を担当するポストフィルタを開発し、IVA と組み合わせました。

1. 空間時変性補償：ロボットが動くことによる空間伝搬モデルの変動を補償するため、時間変動成分を自律推定することの出来る2次統計量制御型・時変フィルタを挿入しました。
2. 統計的音声強調：IVA では達成できない不要音抑圧能力を補償するため、人間の聴覚印象をある程度考慮（高次統計量を制御）した統計的音声強調処理を挿入しました。

上記2つの拡張により、オペレータ等の要求に応じて、「雑音を強力的に抑圧してとにかく音声らしきものを見つける」とことと「被災者が何を言っているのか明瞭化する」とことの両立を図ることが可能となります。また、本ポストフィルタ処理系は、手法 A と併用することも可能です。

<今後の展開>

倒壊家屋内での音声聞き取りにおいて高い性能を発揮できることから、今後、災害救助システムとしての実用化を進めていきます。また、瓦礫内の搜索性能（ロボットの運動能力、および、聞き取り能力）のさらなる向上を図っていきます。

本成果は、以下の事業・研究プロジェクトによって得られました。

- 内閣府 革新的研究開発推進プログラム (ImPACT)

<http://www.jst.go.jp/impact/>

プログラム・マネージャー：田所 諭

研究開発プログラム：タフ・ロボティクス・チャレンジ

研究開発課題 1：索状ロボット（細径）の研究開発

（研究開発責任者：田所 諭、研究期間：平成26年度～平成30年度）

研究開発課題 2：極限音響の基礎技術の研究開発

（研究開発責任者：奥乃 博、研究期間：平成26年度～平成30年度）

研究開発課題 3：ブラインド音源分離と高次統計量制御型音声強調に基づく柔軟ロボット音響センシング

（研究開発責任者：猿渡 洋、研究期間：平成27年度～平成29年



■田所 諭 ImPACTプログラム・マネージャーのコメント ■



ImPACTタフ・ロボティクス・チャレンジは、災害の予防・緊急対応・復旧、人命救助、人道貢献のためのロボットに必要な不可欠な、「タフで、へこたれない」さまざまな技術を創りだし、防災における社会的イノベーションとともに、新事業創出による産業的イノベーションを興すことを目的とし、プロジェクト研究開発を推進しています。

これまで、瓦礫内の要救助者を発見することを目的として能動スコープカメラが開発されてきました。音声は要救助者を発見するための有力な手がかりですが、従来の能動スコープカメラのマイクロホンでは、ロボットの運動が発生するノイズや外部からの騒音にかき消され、瓦礫の遮音性が高いこともあつ

て、要救助者からの声を聞き取ることが困難でした。

本研究は、高い瓦礫内進入性能を持つ能動スコープカメラの新しい設計と多数のマイクロホンの搭載、そのシステムの特長を活かした音声抽出・強調の新しい技術の開発、により、この課題に対する非連続イノベーションを達成したところに意義があります。

今後さらなる実用化を進めることによって、地震災害の瓦礫等における人命救助の飛躍的向上を図るとともに、設備診断などの産業分野への広い波及効果が期待されます。

<用語解説>

注1) マイクアレイ処理

複数のマイクを使用して、音の到来方向を知る音源定位、混合音に含まれる音を分離する音源分離という処理から、さらに進んで、分離音の認識、例えば音声認識や鳥の鳴き声認識などを行う技術。

ロボットに装着したマイクではロボットの動作音や自分の発話、環境音などいろいろな音が混ざった混合音を処理する必要があり、マイクアレイ処理は不可欠になる。

注2) スペクトログラム上のスパース性

入力音の振幅スペクトログラム中の大部分の成分は雑音成分であるのに対し、音声成分はある範囲に飛び飛びの値しか存在しない。この音声成分の疎（大部分が零の値であるが時々大きな値を

とること)な構造を「スパース」と呼ぶ。一方で、ロボット等で発生する雑音は振幅スペクトログラム上で繰り返し構造を持つことが多く、この状態を時間・周波数の2次元関数を行列に見立てて「低ランク」構造と呼ぶ。よって、音声をスパース、ロボット雑音を低ランクで近似すれば音源分離が出来ることとなる。

注3)「NMF (非負値行列因子分解)」

振幅スペクトログラムは、その要素が全て非負の2次元(時間・周波数)関数として表現され、それを比較的少ない本数の基底ベクトル(典型的な周波数特性を表す)とそのアクティベーション

(時間変化を表す)の積に分解表現する技術をNMFと呼ぶ。NMFを使うことにより、音響信号をより簡潔かつ操作しやすい形で表現することが可能となる。

注4) 振幅スペクトログラム領域

入力音に対して短いフーリエ変換を連続的に施したものをスペクトログラムといい、音を「時間」・「周波数」の2次元関数として表すことができる。このスペクトログラムの絶対値をとったものを振幅スペクトログラムと呼ぶ。

注5) IVA (独立ベクトル分析)

ある音源のスペクトログラムを1ランクのベクトルの時間変動とみなし、その各要素が複数音源間で独立であると仮定する音源分離手法。本技術により、観測信号のみから独立な挙動を示す複数の音源を分離抽出することが可能となる。

注6) 線形時不変フィルタ

入力と出力の関係が線形(信号の定数倍や和に関してその性質が保たれること)であるフィルタにおいて、そのフィルタ特性が時間変動を伴わないもの。例えば、ロボットがほぼ移動しない、または非常にゆっくりと移動している場合には、人間の声が空中を伝搬してマイクへ入力される状態は近似的に「線形時不変フィルタ」で表すことが可能となる。