

NEC の音環境理解技術と 国際技術評価イベント DCASE2016 への取り組み

近藤玲史^{†1} 小松達也^{†1} 戸泉貴裕^{†1} 加藤正徳^{†1} 仙田裕三^{†1}

概要: NEC における音環境理解技術への取り組みについて紹介する。音響監視を中心に、技術要素としての音響イベント検知・音源方向推定技術を概説する。企業における研究開発活動で重要な実証活動と技術評価活動について、特に後者で DCASE を活用した事例を紹介する。

キーワード: 音環境理解, 音響監視, 音響イベント検知, 音源方向推定

NEC's Acoustic Environment Understanding Technology and International Technology Evaluation Event "DCASE2016"

REISHI KONDO^{†1} TATSUYA KOMATSU^{†1} TAKAHIRO TOIZUMI^{†1}
MASANORI KATO^{†1} YUZO SENDA^{†1}

Abstract: Introducing NEC's acoustic environment understanding technology. Especially in the area of acoustic surveillance, describing technologies such as acoustic event detection and acoustic direction estimation. Also describing our Proof of Concept activity and technology evaluation activity at DCASE, that are important in company R&D.

Keywords: Acoustic environment understanding, Acoustic surveillance, Acoustic event detection, Acoustic direction estimation

1. はじめに

NEC における音環境理解技術への取り組みについて紹介する。ここで音環境理解技術とは音響信号により周囲の広い範囲（環境）の状況を理解することを狙ったもので、見守り、計測、インシデント検知など広い範囲の用途を想定している。

NEC では、音声符号化[1][2]（製品としては携帯電話、公共無線向けコーデックなど）、音声認識[3][4][5][6][7][8][9]（VoiceGraphy, VoiceDo など）、音声合成[10][11]（VoiceOperator/音声合成, VoiceDo など）オーディオ符号化[12][13][14]（地上デジタル放送向けコーデック, Silicon Audio など）、雑音抑圧[15][16][17]（EuphoMagic など）、耳音響認証[18][19] など音に関する多くの技術の研究開発と製品化を行ってきた。こうした技術を発展し、音環境理解の技術開発に取り組んでいる。

2. 音響監視

音環境理解技術の一つの応用として、音響によって事象（インシデント）を監視し、音の発生、時刻、位置などを警備員や管理者に報知する音響監視がある。監視対象となる事象は危険そのものあるいは危険につながるものが主で、我々は悲鳴や叫びに代表される人間の発する音、呻き声な

ど動物の発する音、破壊音や機器故障による異常音など物体起因の音など多岐に渡る音を対象とし、観測信号中に含まれるそれら目的音の検知、識別、そして発生位置の推定を行う技術を開発している。

以下では、音響監視で必要とされる二つの主な技術、音響イベント検知技術と音源方向推定技術について簡単に説明する。

2.1 音響イベント検知技術

音響イベント検知技術は、音響信号の発生原因となる物理現象や音源などを認識/識別する技術であり、検知したい特定の音を音響イベントと称する。音響信号は物体自体が発するため画像のような光源を別途必要とせず、障害物を回り込んで伝搬する特性もあるため、音響監視の用途では音響イベントの検知によるメリットは大きい。

反面、後述するように実環境ではさまざまな雑音や環境音が継続的または偶発的に発生している。その中から音響イベントを検知することは容易とは言いがたく、言い換えると、実環境における音響イベント検知の課題は雑音や環境音をはじめとした非目的音との戦いとなる。そこで我々は、あらかじめ収集した音響イベントデータをもちいた教師有り音源分離手法により観測信号を複数の音の成分に分解し信号中における該当音響イベント成分の有無を判別する手法により、非目的音混在への耐性を高めた。さらに学習データにない未知の雑音を分解と同時に学習することで実環境における未知の雑音への頑健性を上げる半教師有り

^{†1} NEC データサイエンス研究所
Data Science Research Laboratories, NEC Corporation

手法を開発した。

観測信号を複数の音の成分に分解する方法としては、従来から主成分分析 (PCA), 独立成分分析 (ICA), 非負値行列因子分解 (NMF) など多数の手法が試みられている。これらは何を基準に音の成分を分割するかの違いであり、我々は一般の音響イベントを対象にする上で信号の特性に依存した分解は望ましくないこと、また実用上は単一音響センサー(マイクロフォン)での検知が望ましいことから、NMF を採用している。

NMF により音響信号のスペクトルを各音源スペクトルの非負和としてモデル化することで、観測された音響信号を各基底の重ね合わせへと分解する。ところが従来手法では各基底同士の関係には制約が存在しなかったため、学習した時点で基底の張る特徴量空間が重なったり広過ぎる空間を張る基底、言い換えれば分離性能の悪い基底の組が生成されてしまう問題があった(図1)。そこで我々は、Mixtures of Local Dictionaries (MLD) の考え方を適用し、こうした基底を予め部分集合(基底グループ)に分割することで、音響特徴量空間を適切に分割する方法を考案した(図2)[20]。

MLD により基底の張る空間が適切に制御されることで、それぞれの基底グループによって検知される音響特徴量は、音の細かな種類に対応する。検知したい音響イベントに限らず分離したい音響事象全体を見ると、音は細かな音の組み合わせでできていることが容易に類推でき、かつそれらの元になっている細かな音は複数の音響事象の間で重なりがある。そこで、各基底グループの張る空間上の各基底は全て代表ベクトルで代表するアクティベーションアグリゲーションにより簡約化する。最終的には、細かな音の集合としてサポートベクトルマシン (SVM) によりイベントの有無を検知する(図3)。

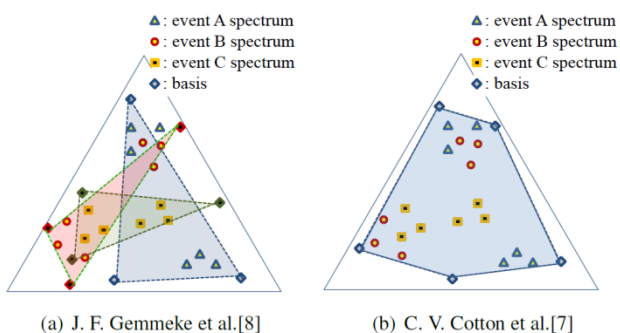


図 1 従来の NMF による基底の張る空間の例(simplex 面上)

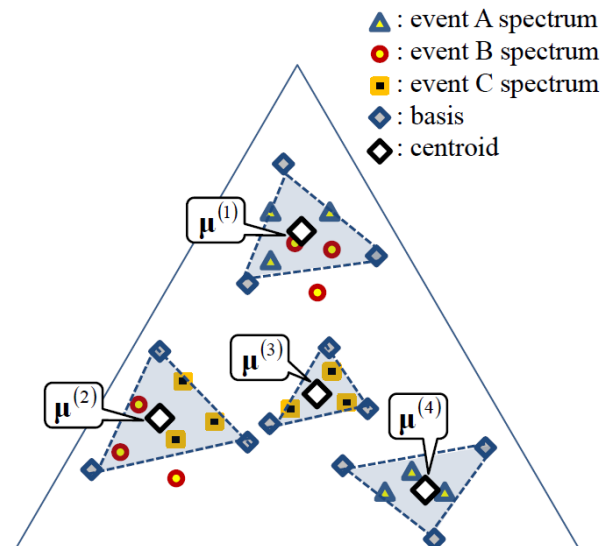


図 2 本手法による基底の張る空間とアクティベーションアグリゲーション

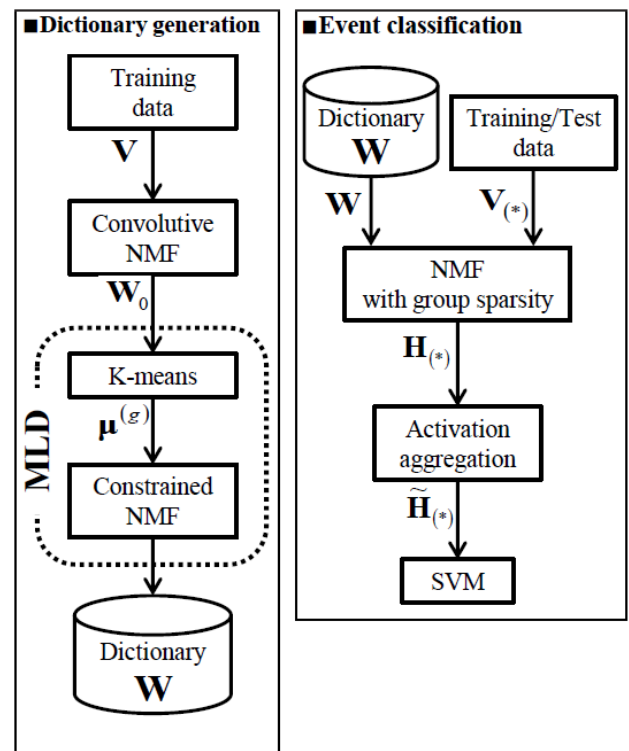


図 3 MLD-NMF とアクティベーションアグリゲーションによる音響イベント検知処理

ここまでで、観測信号を複数の音のモデルの成分に分解することで対象の音響イベントを検知する手法を説明したが、現実の観測信号にはモデルに属さない未知の雑音成分も存在し、それが音源分離を妨げる原因となる。半教師有り NMF を適用して雑音に相当する基底を検知時に生成することで、雑音成分に頑健な音源分離を実現する(図4)[21]。

2.2 音源方向推定技術

音響監視を行う上で、音の発生、時刻に加えて、音源の位置情報が必要になるケースがある。例えば、悲鳴や銃声などの危険音を検知した際に、監視カメラを発生方向に向けた上で、係員に通報することが要請される。監視範囲全域にわたって音響センサーを稠密に配置することは多くの場合困難であるから、局所的（あるいは疎）に配置された音響センサーから音響イベントの方向（位置）を推定することも重要である。

音源方向の推定には複数マイクロフォンへの到達時刻の差から方向を推定する手法がよく知られているが、たとえば屋内での使用を考えると、反射音や雑音環境による精度の低下が問題となる。そこで我々は、二つの音響信号間の相互相関関数を算出してオンセットを高精度に検知、その時刻周辺の相互相関関数に対して先鋭度で重みづけることにより、直接音の音源方向を高精度に推定する技術を開発した[22]。

3. 実証活動

研究開発成果の効率かつ迅速な事業化は、近年の企業 R&D にとって特に重要な課題である。そのためには技術を深耕するに留まらず、実際に技術を必要とする場で運用し、課題を発掘して改良を繰り返す技術開発が効果的と考えている。

NEC は「価値共創研究所」のコンセプトのもと、日本、北米、欧州、中国、シンガポールの世界 5 地域に研究開発拠点を開設している。また世界各地の事業拠点には、個別のお客様やそれぞれの地域によって異なる、あるいは共通する課題が存在する。そうした課題に対して、お客様とともに技術を価値に変換する実証実験を設計・運用することで先進技術の適用や課題の掘り起し、ひいては社会価値としてグローバルに展開を進めていく活動として実証活動を進めている。

一つの例として、シンガポール政府は真に役立つ監視ソリューションを作るため、実際の環境での数ヶ月におよぶ Singapore Safe City Test Bed Project と呼ばれる実証実験(技術試験)を計画した [3]。内務省および経済開発庁が主導する本実証実験プロジェクトには、警察や陸上交通庁をはじめとする複数の政府機関が参加し、喧嘩、うるつきなどの不審行動、異常混雑や放置された不審物など、様々な異常事態の自動検知が求められた。NEC はその一つに参加し、多数の人々が行きかう屋内の公共スペースにおいて、音響・映像解析技術により異常事態を自動検知する技術実証実験を完遂した。

別の例として、ニュージーランドの首都ウェリントン市

における Smart capital に向けた活動がある [3][23]。同市はスマートシティを積極的に推進しており、NEC は Wellington City Council と共に同市内の各所で実証実験を行っている。同市の取り組みは通行量測定、オープンデータ化など多岐に渡るが、音響検知ではガラスの割れた音や叫び声が上がったことの検知が含まれる。こうした活動では、プライバシーやセキュリティも重要な要素となっている。

4. 技術評価活動

研究開発成果を強化し事業化していく上では、技術の相対的な強さを客観的に評価することも必要である。前述の実証活動では特定の対象・状況に対する技術の強さを示すことができるが、一般的には他者との比較は困難なことが多い。共通のデータ・スキームを用いた音響監視に関係する技術評価としては、過去には CLEAR (2006, 2007), D-CASE (2013) などの競争型ワークショップが開催されてきた[24]。

今春、音響シーンとイベントの検知・分類を対象とした競争的国際技術評価イベント“Detection and Classification of Acoustic Scenes and Events 2016 (DCASE2016)”が IEEE AASP Challenge の一つとして開催された [25]。我々は task 2 (Sound event detection in synthetic audio) に前述の音響イベント検知方式(図 4) [21] を用いて参加し、公開された提出テストデータセットに対する評価結果で、参加 10 組織中 overall ほかで最も良い成績を獲得することができた。これにより、技術の強さの相対的な位置づけを確認した。

今回定義された 4 種類のタスクの中で task 2 には、対象とする音イベントの幾つかは実際の音響監視で現れる検知対象音響イベントや検知の妨害となる環境音に近い特性のものが含まれている。先に述べたように、我々は音響イベント検知として観測信号から音を分離するアプローチを採っており、同タスクで高い性能を示すことにより分離性能の高さと音響監視に向けた有用性を示すことができた。

あわせて公開されている開発データセットを用いてセグメントベースの F 値を評価したところ、SNR の異なる各データセットでの比較において同 F 値の低下をベースラインに比べて低減できていることを確認できた(図 5)。これは開発した技術が雑音に頑健であることを示しており、実際の音響監視への応用を考えると大きなメリットと言える。

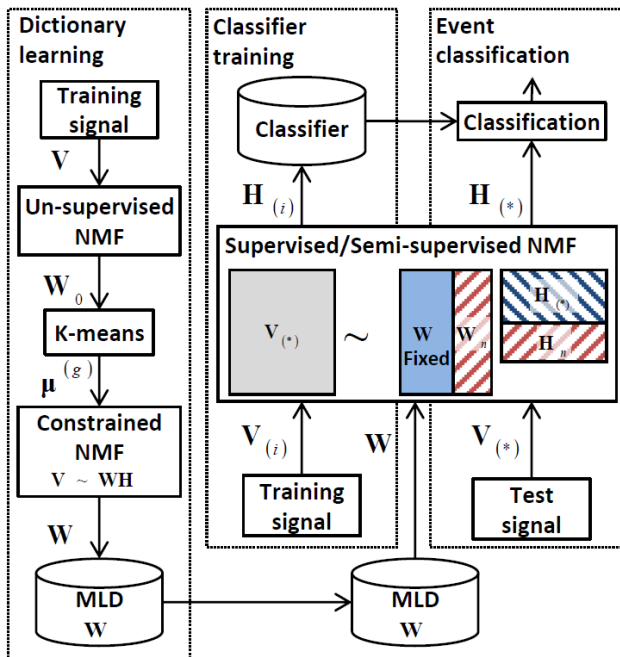


図 4 DCASE2016 参加時の方式概要

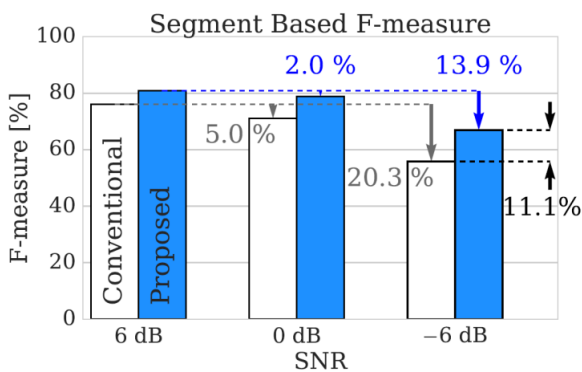


図 5 DCASE2016 システムによる
 SNR ごとの F 値低下割合

5. まとめ

NEC における音環境理解技術への取り組みについて紹介した。音響監視を中心に、技術要素としての音響イベント検知技術、音源方向推定技術を概説した。また企業における研究開発で大切となる実証活動と、技術評価活動の取り組みについて紹介した。今回技術評価活動では、競争的国際技術評価イベント DCASE2016 task 2 に参加し、参加組織中もっとも良い成績を獲得、技術の強さの相対的な位置づけを確認した。

参考文献

[1]小澤一範, 田海真一, 野村俊之, "マルチパルスベクトル量子化音源と高速探索に基づく MP - CELP 音声符号化", 電子情報通信学会論文誌 A, Vol. J79-A, No.10, pp.1655-1663, Oct., 1996.
 [2]野村俊之, 岩垂正宏, 田中直也, "MPEG-4/CELP 音声符号化方式", 電子情報通信学会技術研究報告. SP, 音声 98(424), 19-26, 1998-11-20
 [3]谷真宏, 仙田裕三, 近藤 玲史, 越仲 孝文 " NEC シンガポール

研究所と音声・音響解析への取り組み", SLP-108
 [4]寺西博人 " 音声認識技術活用 議事録作成支援ソリューション「VoiceGraphy (ボイスグラフィー)」 2013 年度 グッドデザイン賞受賞 ", SLP-103 (2014)
 [5]服部浩明 " 耐雑音音声認識装置 VoiceDo の応用 ", SLP-98 (2003)
 [6]花沢健, 辻川剛範 " キャラクターとの会話体験を提供する音声応答の試験サービス ", SLP-93 (2012)
 [7]寺西博人 " 音声認識ソリューション開発について ", SLP-68 (2007)
 [8]越仲孝文, 江森正, 大西祥史 " 裁判員裁判向け音声認識システム ", NEC 技報 Vol.63 No.1 (2010 年 2 月)
 [9]越仲孝文, 宝珠山治, 大西祥史, 磯谷亮輔, 谷真宏 " 音声・音響分析技術とパブリックソリューションへの応用 ", NEC 技報 Vol.67 No.1 (2014 年 11 月)
 [10]加藤正徳, 近藤玲史, 三井康行 " 電話応答サービスに適した音声合成の開発 ", SLP-88 (2011)
 [11]三井康行 " 音声合成の利用シーンと要求される品質との関係 ", SLP-83 (2010)
 [12]A. Sugiyama, F. Hazu, M. Iwadare and T. Nishitani, " Adaptive Transform Coding with an Adaptive Block Size (ATC-ABS), " Proceedings of ICASSP'90, pp.1093-1096, Apr. 1990.
 [13] A. Sugiyama, M. Iwadare, N. Ohdate, T. Manabe, H. Takano, O. Kitabatake and E. Hirao, " The Silicon Audio --- An Audio-Data Compression and Storage System with a Semiconductor Memory Card ---, " IEEE Transactions on Consumer Electronics, Vol.41, No.1, pp.186-194, Feb. 1995.
 [14] 野村俊之, 嶋田修, 高見沢雄一郎, 芹沢昌弘, 田中直也, 津島峰生, 則松武志, " MPEG-4 オーディオ標準化向け低演算量帯域拡張方式 ", 電子情報通信学会総合大会講演論文集, D-14-8 (2003 年)
 [15]M. Kato, A. Sugiyama and M. Serizawa, " Noise suppression with high speech quality based on weighted noise estimation and MMSE STSA, " Proc. IWAENC2001, pp. 183-186, Sep. 2001.
 [16]M. Kato and A. Sugiyama, " A Wind-Noise Suppressor Based on Wind-Onset Detection and Spectral Gain Modification, " Proc. IWAENC2014, pp. 145-149, Sep. 2014
 [17]A. Sugiyama and R. Miyahara, " A Directional Noise Suppressor with a Specified Beamwidth, " Proc. ICASSP2015, pp. 524-528, Apr. 2015
 [18]荒川隆行, 矢野昌平, 越仲孝文, 今岡仁, 入澤英毅 " 外耳道音響特性を用いた高精度個人認証 ", 日本音響学会 2016 年春季研究発表会講演論文集 2-P-30 (2016 年 3 月)
 [19]T. Arakawa, T. Koshinaka, S. Yano, H. Irisawa, R. Miyahara, and H. Imaoka " Fast and Accurate Personal Authentication Using Ear Acoustics, " APSIPA 2016 (To be published in December)
 [20]Tatsuya Komatsu, Yuzo Senda, Reishi Kondo " Acoustic Event Detection Based on Non-negative Matrix Factorization with Mixtures of Local Dictionaries and Activation Aggregation, " ICASSP2016
 [21]Tatsuya Komatsu, Takahiro Toizumi, Reishi Kondo, Yuzo Senda " Acoustic Event Detection Method Using Semi-supervised Non-negative Matrix Factorization with a Mixture of Local Dictionaries, " DCASE2016 (2016.6)
 [22]荒井友督, 仙田裕三 " 複数マイク信号間の相互相関関数に基づく残響環境下の音響オンセット検知と方向推定 ", 電子情報通信学会 第 30 回信号処理シンポジウム P1-16 (2015 年 11 月)
 [23] " 'Smelling' graffiti: Wellington trials new smart city technology, " The National Business Review, p.18,20, September 25, 2015
 [24]大石康智 " あらゆる音の検出・識別を目指して : 音響イベント検出研究の現在と未来 " 日本音響学会研究発表会講演論文集 1521-1524 (2014)
 [25] " Detection and Classification of Acoustic Scenes and Events 2016, IEEE AASP Challenge, " <http://www.cs.tut.fi/sgn/arg/dccase2016/> (参照 2016 年 9 月 5 日)